Technical Report 528

COST AND INFORMATION EFFECTIVENESS ANALYSIS (CIEA): A METHODOLOGY FOR EVALUATING A TRAINING DEVICE OPERATIONAL READINESS ASSESSMENT CAPABILITY (DORAC)

John K. Hawley and Edward D. Dawdy
Applied Science Associates, Inc.

ARI FIELD UNIT AT FORT BENNING, GEORGIA

U. S. Army
Research Institute for the Behavioral and Social Sciences

February 1981

Approved for public release, distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

FRANKLIN A. HART
Colonel, US Army
Commander

Research accomplished under contract
for the Department of the Army

Applied Science Associates, Inc.

NOTICES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report 523	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COST AND INFORMATION EFFECTIVENESS ANALYSIS (CIEA): A METHODOLOGY FOR EVALUATING A TRAINING DEVICE OPERATIONAL READINESS ASSESSMENT CAPABILITY (DORAC)		5. TYPE OF REPORT & PERIOD COVERED Final Report Dec 1979 - Feb 1981 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John K. Hawley and Edward D. Dawdy		8. CONTRACT OR GRANT NUMBER(s) MDA903-80-C-0212
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Science Associates, Inc. P.O. Box 158 Valencia, PA 16059		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263743A794, 334353
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Ave., Alexandria, VA 22333		12. REPORT DATE February 1981 13. NUMBER OF PAGES 68
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) ARI Field Unit P.O. Box 2086 Ft Benning, GA 31905		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Training Device	Multiattribute Utility Measurement
Operational Readiness	Cost-Effectiveness Analysis
Information	Performance Evaluation
Information Value	Performance Measurement
Information Utility	Information Amount

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents an overview of a problem that accompanies the development of a Training Device Operational Readiness Assessment Capability (DORAC): determining when the value of the information received from the DORAC is greater than or equal to the cost of obtaining it. This determination may be needed as a part of specifying the requirements for a device, or as a basis for deciding between two or more design options developed to satisfy the requirements. An analysis focused on this issue is termed Cost and Information Effectiveness Analysis (CIEA).

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19.

Performance Assessment	Sensitivity Analysis	Partial Utility
Utility Theory	Information Needs	Aggregate Utility
Measurement	Worth Dimensions	Additive Model
Psychological Scaling	Information Measures	Multiplicative Model
Level of Measurement	Problem Structuring	Information Quality
CIEA	DORAC	

20.

As a first step, the concept of information is defined and characterized by two attributes: Amount and value. Amount is measured formally through indices like Shannon's H; value is a function of the changes resulting from the receipt of information by decision-makers.

Because of the problems inherent in determining strict information value, it is proposed that information utility be used as a proxy measure for information value. Accordingly, an information worth assessment procedure based on multiattribute utility measurement (MAUM) is developed and presented. The MAUM information worth assessment procedure is integrated into a standard cost-effectiveness analysis methodology to form the core of a preliminary CIEA methodology, denoted MAUM-CIEA. An exemplary CIEA for a hypothetical set of DORAC alternatives is presented with the methodological discussion.

The final section of the report summarizes the assumptions underlying the use of MAUM-CIEA. In addition, methodological issues relevant to the widespread application of MAUM-CIEA are listed and discussed. Recommendations for additional methodological development are also presented.

Technical Report 528

COST AND INFORMATION EFFECTIVENESS ANALYSIS (CIEA): A METHODOLOGY FOR EVALUATING A TRAINING DEVICE OPERATIONAL READINESS ASSESSMENT CAPABILITY (DORAC)

John K. Hawley and Edward D. Dawdy
Applied Science Associates, Inc.

Submitted by:
H.C. Strasel, Chief
ARI FIELD UNIT AT FORT BENNING, GEORGIA

Approved by:
E. Ralph Dusek, Director
TRAINING RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

February 1981

Army Project Number
2Q263743A794

Assessing Device Operational
Readiness Assessment Capability (DORAC)

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The research reported here is part of a broader program on training for combat effectiveness being conducted by the US Army Research Institute for the Behavioral and Social Sciences. The availability of current and sufficient knowledge of troop combat readiness and existing skill deficiencies is critical to resource allocation and training management in units. The ARI Field Unit at Fort Benning, Georgia, is developing the methods and guidelines necessary to implement a partial answer to this need: the design and use of unit training devices to satisfy personnel evaluation and qualification purposes also. This report describes an analytical method for assessing the cost effectiveness of using a particular device or any one of a set of alternative devices, and how often to use it, for evaluation of personnel combat readiness and skill deficiency.

The method will be useful to training developers (e.g. USAIS DTD) in deciding whether or not to require a capability on the part of a training device to evaluate personnel performance deficiencies and qualifications. The method will be useful to device developers (e.g. PM TRADE) in deciding which training device would best implement the requirement from both cost and benefit standpoints.

JOSEPH ZEIDNER
Technical Director

COST AND INFORMATION EFFECTIVENESS ANALYSIS (CIEA): A METHODOLOGY FOR EVALUATING TRAINING DEVICE OPERATIONAL READINESS ASSESSMENT CAPABILITY (DORAC)

BRIEF

Requirement:

Up to date information on the combat readiness and skill deficiencies of troops is needed to report personnel readiness levels and to manage troop training. Meeting this need can be made difficult by frequent personnel turnover; infrequent testing; and unknowns regarding skill decay rates. A partial solution to this problem may reside in unit training devices that are currently fielded and/or under development. If a device is designed and used to satisfy training requirements, it may, in many cases, also be designed and used on a regular basis (e.g., quarterly) to determine personnel qualification status and skill deficiencies requiring training. However, the inclusion of the evaluation capability may, sometimes, be more expensive than the payoff can justify. This calls for a cost benefits analysis method to assess whether a particular device, or one of a set of device design alternatives, is worth using in this manner. I.e., a method to determine if the information that can be gained from evaluation of performance on the device is worth the costs of the evaluation process and any necessary modifications to the device design (e.g., incorporation of an expanded visual simulation capability or a performance measurement software system).

Procedure:

A review of the literature on cost benefits, information theory, and utility evaluation was conducted to identify and examine current methods. Existing models and techniques were integrated and further developed into a multiattribute utility measurement (MAUM) Cost and Information Effectiveness Analysis (CIEA) method. The method was tested and further refined through application to a set of testing alternatives for M16A1 rifle marksmanship skills. The alternatives included the Weaponeer, Squad Weapon Analytical Trainer (SWAT), and the current field testing procedure, Record Fire. This report presents the details of the developed MAUM CIEA method and presents fictitious data for illustration purposes. The details of the application and actual resulting data are presented in a companion report (Hawley & Dawdy, 1981).

Findings:

(1) The developed MAUM CIEA procedures were applied successfully in an evaluation of alternatives for testing M16A1 rifle marksmanship skills. Army officers with training device expertise, and other officers with operational unit responsibilities, were able to apply the MAUM CIEA to the evaluation and the resulting values appeared reasonable.

(2) There is a need, however, to simplify and expedite the application of the MAUM CIEA method to the extent possible. Possible avenues of approach will be explored in future research.

Utilization of Findings:

The ultimate users of the MAUM CIEA method will be US Army Training and Doctrine Command (TRADOC) training developers and device design engineers working with the materiel developer, PM TRADE. The method will be used to determine if and how to develop and implement DORAC in existing training devices and in devices under development. Further work will be done to simplify and validate the method to make it practical for use by these personnel.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SECTION 1

INTRODUCTION

Background

Most scenarios for a full-scale confrontation between the United States and any of its major potential adversaries indicate that the majority of Army units will have to be prepared to fight immediately without the luxury of a lengthy mobilization period, such as the first year of World War II. Studies of the comparative military strengths of the United States and the Warsaw Pact countries also indicate that U.S. forces are likely to be heavily outnumbered, often by a ratio of five to one or more. To have any hope of success in this "come as you are and win the first battle outnumbered" situation, the Army will have to maintain a high level of individual and unit combat readiness at all times. Maintaining a consistently high level of combat readiness will necessitate frequent evaluations of individual and unit proficiency, along with a means of quickly diagnosing and remediating performance problems.

In simpler times, the assessment of individual and unit job proficiency presented no special difficulties. Recently, however, the complex nature of many weapons systems, personnel turnover rates, and changes in training philosophy have led to an emphasis on performance-oriented training and criterion-referenced testing. Soldiers are required to demonstrate individual and collective competencies in a "hands on" environment using actual equipment. This change in training and evaluation methodology has increased the demand for training/evaluation uses of operational equipment and accompanying support resources requirements [e.g., ammunition, spare parts, POL (petroleum, oil, and lubricants), etc.] during a time of inflation and budgetary constraints. More frequent performance-oriented individual or collective readiness evaluations would tend to complicate this already tight situation.

A proposed solution to the problem of conducting more frequent readiness evaluations in the face of tight resource constraints is to use training devices instead of actual equipment in the conduct of these evaluations (Hopkins, 1975). In addition to their uses in training, training devices (e.g., simulators, mock-ups, etc.) can often provide a vehicle for individual or collective proficiency evaluations (Fitzpatrick & Morrison, 1971; Glaser & Klaus, 1972; Finley, Gainer, & Muckler, 1974; Crawford & Brock, 1977). Historically, the most extensive use of training devices in proficiency assessment has been in aviation (Caro, 1973). The commercial airlines and the Federal Aviation Administration use flight simulators extensively in proficiency assessment. Follow-up studies have indicated that pilot performance in flight simulators is predictive of performance in actual aircraft (American Airlines, 1969; Weitzman, Fineberg, Gade, & Compton, 1979).

Within a military setting, the uses of training devices in performance evaluation have generally mirrored civilian uses and primarily involved aviation. There has been, however, an increasing use of training devices to assess individual and collective proficiency in other areas. Among the other applications have been maintenance performance and anti-submarine warfare (ASW) crew proficiency. In the Army, one long-standing program of individual and collective readiness assessment using a training device is found in the Air Defense branch. Here, the AN/TPQ-29 simulator (and prior to that the AN/MPQ-T1 simulator for the Nike-Hercules system) is used in the conduct of operation readiness evaluations for HAWK Air Defense units. The AN/TPQ-29 is an engagement simulator capable of generating a variety of simulated air defense combat situations [e.g., multiple targets, electronic countermeasures (ECM) of various kinds, etc.]. The simulator was designed primarily for use as a training device, but it can be (and is) used to evaluate individual and crew performance. When using the AN/TPQ-29 in assessment, an evaluation team loads a "raid tape" containing the parameters of an air defense engagement into the simulator. The HAWK crew is evaluated on its ability to defeat the simulated threat; performance checklists are used to evaluate individual crew member performances. Hardcopy printouts of some individual and crew performance measures (e.g., targets destroyed, numbers of penetrators, engagement times, operator reaction times, etc.) are also obtainable from the simulator.

The evaluation of HAWK personnel using the AN/TPQ-29 engagement simulator illustrates the concept of a Training <u>D</u>evice <u>O</u>perational <u>R</u>eadiness <u>A</u>ssessment <u>C</u>apability (DORAC). DORAC simply means that an operational readiness assessment capability is included with the training devices for a materiel system. Once built into the training device system, the DORAC can be used to assess the operational readiness of individuals or crews. As an added feature, the measurement capability inherent in a DORAC can also provide information useful for other purposes such as training management, unit management, materiel system evaluation, and the like.

A recent review of Army training device proficiency assessment capabilities has indicated that the DORAC principle *can* be applied to the training devices for virtually any parent materiel system (Shelnutt, Smillie, & Bercos, 1978). At the present time, actual use as in the HAWK system and in the aviation community is somewhat rare, but the potential remains. There are, however, several issues that remain to be resolved before generally attempting to implement the DORAC principle within the Army. Among these issues are the feasibility of the concept, the utility of the resulting information, and a means for determining what performance data are actually required. One of the primary issues involves a workable methodology for establishing the cost-effectiveness of a DORAC; that is, determining the conditions under which the information (readiness, training status, etc.) derived from a particular DORAC is worth the cost required to obtain it. This determination may be needed as a part of specifying the requirements for a device or as a basis for deciding between two or more design options developed to satisfy the requirements. The objective of this report is to present a methodology, denoted Cost and Information Effectiveness Analysis (CIEA), for evaluating DORAC alternatives in terms of the value of the information provided versus the cost of developing, implementing, and operating the capability. The alternatives include a baseline case, i.e., the current operational readiness procedure and frequency (e.g., an annual ARTEP); and one or more training device designs operated at one or more annual frequencies for evaluation purposes. The costs of developing and operating the device to satisfy training requirements are not included in the DORAC CIEA; the only costs included are those associated with a simulation or measurement capability, or evaluation processes (e.g., administration support costs), which are over and above those required to support training requirements per se. CIEA is intended to serve as a guide for decision-makers in deciding whether DORAC should be included in any training device, irrespective of the parent materiel system.

Assessing Information Worth

The ultimate objective of CIEA is to facilitate assessing the relative worth of the differential information obtained from alternative DORACs that have different costs and other associated resource requirements. Information (readiness, training status, etc.) has value only when it results in a gain to the receiving party; thus, information should be collected only when the incremental cost of obtaining it is less than the gain that is realized through its receipt. CIEA is a procedure for determining whether or which DORAC alternatives are worth implementing. Estimating the cost required to obtain a given amount of information is a relatively straightforward procedure. However, determining the value of the same quantity of information is not as straightforward. A given piece of information may have value for one purpose but not for another. Information value depends on a particular communication's relevance for potential action by a specific set of decision makers acting in a specified environment at a certain point in time.

Before addressing the problem of assessing information worth, it is first necessary to define what is meant by the term information. Bedford and Onsi (1966) define *information* as "data evaluated for a specific use." Reviewing information is characterized as a process of ignorance reduction. The function of information is to reduce the amount and range of uncertainty under which decisions are made.

Two attributes are associated with information: *Amount* and *Value*. The amount of information in a communication is determined by the reduction in uncertainty resulting from its receipt. Amount is assessed formally through the application of measures such as Shannon's H (Shannon, 1948;

Shannon & Weaver, 1963). Information value is a function of the *changes* resulting from the use of information in the pursuit of a particular purpose (Bedford & Onsi, 1966). A communication may contain a large amount of information without being valuable in the sense of saying something useful to the recipient. Information value is measured by a receiver in terms of the information's uses in decision making. Strictly speaking, this definition implies that information value is measured by comparing the actions of decision-makers before and after the receipt of a given quantity of information. Value is determined from the gain realized as the result of taking one course of action versus another after receiving a given amount of information (Bedford & Onsi, 1966; Thiel, 1967; Lev, 1969). In the present situation, this view requires that the value of DORAC information strictly be measured by assessing the gain associated with an information recipient's taking one course of action versus another on the basis of having received some amount of information; for example, by training for task X instead of for task Y, or by training P hours this week instead of Q hours.

Determining information value through a decision-maker action substitution approach as described above would seem to be a relatively straightforward procedure. However, some serious potential problems are apparent. First of all, the logistics of such an information worth evaluation procedure would likely be unmanageable. For example, the alternatives of a range of decision makers would have to be considered, the differential costs of numerous before and after decision scenarios would have to be determined, the costs associated with some action differentials might be difficult or impossible to quantify, and so on. Based on a preliminary review of such potential problems, it is doubtful that an action substitution approach to information worth assessment would prove to be a workable or useful approach to the conduct of CIEA.

An alternative worth evaluation approach that holds promise in the current problem situation is to consider information utility rather than information value in its strict sense. Utility is defined as "Psychological value" (Johnson & Huber, 1977). Information utility is thus defined as "the psychological or subjective value of information in the pursuit of a particular purpose." The use of information utility as a proxy measure for information value assumes that high information utility represents at least a propensity to make decisions as the result of the receipt of some given piece of information. An added assumption is that information value (IV) is a monotonically increasing function of information utility (IU); that is,

$$IV_i = \Phi \ (IU_i) \qquad (1\text{-}1)$$

where $IV_i \leq IV_j$ whenever $IU_i \leq IU_j$.

The use of information utility as a proxy for information value reduces the problem of measuring information value to the task of assessing the utility of specific categories of information for specific users. Since a well developed technology is associated with utility measurement, assessing information utility constitutes a more manageable task than assessing strict information value. The next portion of the report presents an overview of utility theory and utility measurement procedures as they relate to the issue of information worth assessment in CIEA.

Utility Measurement

Theory

The branch of utility theory that is most applicable to the CIEA problem is termed riskless multiattribute utility measurement (MAUM). The basic idea underlying MAUM is quite simple (for example, see Raiffa, 1968; Keeney & Raiffa, 1976): Every outcome of an action has a value or utility on a number of different attributes, dimensions, or factors. The objective of MAUM, in any of its numerous versions, is to determine these values, one factor at a time, and then to combine them across factors using a suitable aggregation rule.

To introduce the notation and concepts of MAUM, let x_i be a particular value of factor $\underline{i}$. Furthermore, let X_i, where $x_i \in X_i$, represent the set of all possible values taken on by factor $\underline{i}$. For example, the i^{th} factor could be "color", which might be one of several attributes considered when buying a car. Let there be a total of $\underline{n}$ factors under consideration (e.g., price, style, economy, etc.). Any given alternative is represented as a particular attribute combination:

$$[x_1, x_2, x_3, \ldots, x_n]. \tag{1-2}$$

The tendency to prefer one alternative over another, say a Brand A to Brand B, is represented by a construct called utility, denoted U. In the MAUM context, it is assumed that the utility of an alternative is a function of the utility of the individual attributes; that is,

$$U = F(x_1, x_2, x_3, \ldots, x_n). \tag{1-3}$$

Furthermore, it is assumed that there exists a utility measure for each value of each factor, denoted $u(x_i)$. What MAUM seeks is a functional relationship among the individual factor utilities,

$$U = F[u(x_1), u(x_2), u(x_3), \ldots, u(x_n)], \tag{1-4}$$

that defines an explicit value structure and can serve as a basis for decisions concerning alternatives.

Application

The application of MAUM in the evaluation of alternatives is carried out in three phases, namely:

1. Problem structuring
2. Utility assessment
3. Utility synthesis

Each of the three phases is discussed in the following paragraphs.

Problem Structuring. Phases two and three of the MAUM process provide the tools for evaluating various alternatives. The tools are of limited value, however, unless they are applied to an appropriately structured problem. Structuring a utility measurement problem is the most critical aspect of MAUM. Problem structuring is also the phase that is least amenable to formal methodological treatment. There are no clear-cut procedures that guide an analyst to a correct problem structure.

Research has shown that the complexity of MAUM applications is inversely related to the effort spent in problem structuring (Edwards, 1971; Keeney & Raiffa, 1976; Johnson & Huber, 1977). As a rule, the better the problem is defined, the simpler the MAUM process becomes, and the more usable and acceptable the resulting solution. In terms of the attributes of a correct problem structure, Johnson and Huber (1977) specify three characteristics, listed as follows:

1. The questions and issues underlying the problem situation are isolated and specified.

2. The context within which utility is to be evaluated is well defined.

3. The attributes or factors that are involved in the analysis are well defined.

Detailed discussions of the heuristics of problem structuring in MAUM are presented in Edwards (1971), Edwards (1976), Keeney and Raiffa (1976), or Johnson and Huber (1977).

Utility Assessment. Once an assessment problem has been adequately structured, the second phase of MAUM involves determining the partial utilities of each alternative on each factor. The procedures that are used to determine individual factor partial utility scores are all direct extensions of psychological scaling methods [see Guilford (1954), Torgerson (1958), Coombs (1964), or Stevens (1975)]. These scaling procedures are classified into five categories, listed as follows:

1. Ranking procedures
2. Category methods
3. Direct methods
4. Gamble methods
5. Indifference methods

Ranking methods require ordering the levels of a factor from most preferred to least preferred. These methods are very easy to use; they also have been found to produce results that are acceptable to decision makers. The resulting scale values are interpreted as ordinal measures of the relative worth of each of the alternatives on the factor in question. Ranking methods are not often used as a sole basis for utility assessment. They are, however, frequently used as a first step in utility evaluation, after which another method (e.g., a Direct method) is used to obtain quantitative (i.e., equal-interval or ratio) scale scores (Johnson & Huber, 1977).

The Category scaling methods represent a methodological variant of the Ranking methods. In these methods, alternatives undergoing evaluation are sorted into discrete categories. The categories are ordered to reflect a continuum of worth for the factor under consideration. Category methods are often easier to use than ranking methods, particularly when evaluating a large number of entities (Torgerson, 1958). They are also subject to the same limitations as the Ranking methods in that the resulting scale values are ordinal measures of relative utility.

Direct scaling methods involve the assignment of numerical scale values to the levels of a factor. Typically, a decision-maker is asked to assign a scale value to each factor level; the scale value reflects that level's utility relative to the other levels of the factor. In MAUM, the most frequently used Direct scaling procedure requires a decision-maker first to anchor extreme factor levels at the extremes of a utility scale, usually a 0-to-100 range. Intermediate alternatives are then scaled between these extreme points (Edwards, 1971).

Although the Direct scaling methods are usually used to assess *subjective* utility, they also can be used to assign utility scores to *objective* dimensions. If a factor can be scored according to an objective criterion (e.g., cost), the resulting scale value can be used, along with utility scores on subjective factors, to determine aggregate utility. The only requirement is that all of the values be made commensurable before individual partial utilities are combined. This is accomplished by converting all scores to a common scale, usually 0-to-100 (Edwards, 1971).

The Gamble scaling methods involve constructing wagers (defined as a situation in which each of a possible set of outcomes can occur with a specified probability) and then varying either factor levels or their associated probabilities until a decision-maker is indifferent between the wager and an alternative "sure thing." Gamble methods are the only utility assessment procedure having a basis in classical utility theory (Keeney & Raiffa, 1976). They are also the only methods that require an explicit assessment of the uncertainty associated with various alternatives. Hence, the Gamble methods are appropriate for MAUM situations involving risk; they are not applicable in riskless situations.

In application, Gamble methods have been found to be cumbersome and time consuming. Raiffa (1968) has referred to this situation as the "bushy mess" problem: in many practical problems, the procedure is often too complex to carry out. Also, in many applications, subjective estimates of the uncertainties involved (i.e., probabilities) have been found to be highly unreliable (Keeney, 1977).

Indifference scaling methods represent a variation on the Gamble methods. In these procedures, a decision-maker is asked to determine indifference points between combinations and levels of factors. For example, a decision-maker may be asked to indicate whether he is indifferent between two pairs of factor level combinations, such as 40% reliability and a repair cost of 30 units versus 50% reliability and a repair cost of 50 units. Indifference methods are the only scaling procedures applicable in situations involving dependent factors. However, methods are available for removing or overcoming factor dependencies [see Dawes & Corrigan (1974) or Einhorn & Hogarth (1975)]. Keeney (1976, p.7) has characterized indifference scaling methods as follows:

> "My experience with (Indifference scaling methods) suggests that such hypothetical judgments are unreliable and unrepresentative of real preferences; worse, they bore untutored decision-makers into rejection of the whole process or acceptance of answers suggested by the sequence of questions rather than answers that reflect their real values, or both."

In a review of a number of representative MAUM applications, Johnson and Huber (1977) conclude that Direct scaling methods are preferred for use in MAUM unless the decision problem involves dependent factors that cannot be orthogonalized or involves uncertainty that must be formally considered. The dominance of Direct scaling procedures in most MAUM situations is judged to be a function of the following points:

1. Direct scaling methods are simple, easy to use, have high face validity, and generate quantitative worth scores.

2. Direct scaling methods are integral to many frequently used systems of worth assessment (for example, see Barish and Kaplan, 1978).

3. The results of utility assessments using Direct scaling methods have been found to be comparable to the results of assessments made using more sophisticated procedures such as one of the Gamble methods (Edwards, 1976).

Table 1-1 presents a summary overview of each of the five scaling methods. The table lists factors that should be considered in the selection of a procedure for a specific MAUM problem.

Utility Synthesis. The third and final phase of the MAUM process involves combining partial utility scores to produce a measure of aggregate or global utility for each alternative. Johnson and Huber (1977) present four basic models for combining partial utilities. These models are listed as follows:

1. Linear additive:

$$U = b_o + \sum_{i=1}^{n} b_i u(x_i)$$

2. Linear additive with interactions:

$$U = b_o + \sum_{i=1}^{n} b_i u(x_i) + \sum\sum_{i \neq j} b_{ij} u(x_i) u(x_j)$$

3. Linear additive with higher order functions:

$$U = b_o + \sum_{i=1}^{n} b_i f[u(x_i)]$$

4. Multiplicative:

$$U = b_o \prod_{i=1}^{n} u(x_i)^{b_i}$$

In the above equations, U represents global or aggregate utility, $u(x_i)$ represents the partial utility of having attribute $\underline{i}$ at level $\underline{x}_i$, and b_i represents the importance of the i^{th} factor to overall utility.

Table 1-1

A Comparison of Utility Assessment Techniques
[Adapted from Johnson and Huber (1977)]

	RANKING METHODS	CATEGORY METHODS	DIRECT METHODS	GAMBLE METHODS	INDIFFERENCE METHODS
Input Factors					
Risk/Uncertainty (Probabilities Required)	No	No	No	Yes	No
Continuous or Discrete Factors	Discrete	Discrete	Either	Either	Continuous
Independent Factors Required	Yes	Yes	Yes	No	No
Monotonic Factors Required	No	No	No	No	Yes
Decision Maker(s)					
Response required	Preference judgments	Categorical judgments	Quantitative judgments	Indifference judgments	Indifference judgments
Task Complexity	Very simple	Very simple	Moderate	Moderate	Simple
Training required	Almost none	Almost none	Moderate	Extensive	Little
Acceptability by Decision Makers	High	High	Moderate	Moderate	Generally Low
Face validity	High	High	High	Low	Moderate
Assessment Time	Very quick	Very quick	Moderate	Slow	Slow
Results					
Data Processing	Very little	Very little	Usually none	Moderate	Moderate to high
Output	Relative worth	Approximate numerical worth	Numerical worth	Numerical worth	Relative/ numerical worth
Response	High, if only a few alternatives	High, if only a few alternatives	Moderate to high	Moderate	Moderate to low

Table 1-1 (Cont'd)

	RANKING METHODS	CATEGORY METHODS	DIRECT METHODS	GAMBLE METHODS	INDIFFERENCE METHODS
General					
Skill and Experience required of analyst	Little	Little	Moderate	High	High
Flexibility of Procedure	High	High	High	Low	Moderate

Model 1, linear additive, is the simplest utility aggregation model. Here, the $u(x_i)$ are the simple partial utility scores for each of the factors. The second model, linear additive with interactions, is a more complex version of model 1. In this model, interactions between various factor combinations are also considered. Model 3, linear additive with higher order functions, permits the inclusion of complex individual factor utility functions [e.g., $u(x_i)^2$, $\log u(x_i)$, etc.]. This model is also applicable to situations in which individual factors are multi-faceted; that is, when individual partial utility scores are themselves multiattribute utility scores. Model 4, multiplicative, represents the case of purely interactive utility. Multiplicative models are considerably more complicated to process than the additive models. Also, models of this kind are highly sensitive to changes in the partial utility ratings of individual factors.

All of the additive models (Models 1 through 3) are compensatory; that is, high partial utility scores on one or more factors can compensate for low ratings on other factors. The multiplicative model is far less compensatory than the additive models in that high scores on one or more factors do not as readily compensate for low scores on other factors.

In practical applications, additive aggregation models have been found to be more useful than more complex models (Johnson & Huber, 1977). This conclusion is based on the fact that additive models are simple, quite direct in their application, and allow decision-makers a perceived high degree of control over the MAUM process. John and Edwards (1978, p.1) provide the following comment on the choice of a simple (i.e., not involving interactions) additive model in MAUM: "Most choice situations are non-additive, risky, or both. But more complicated models, involving more complex function forms and more parameters, are often not worth the effort. In particular, the (simple) additive model serves as a good approximation to much more complicated forms (Dawes, 1971; Goldberg, 1965, 1968, 1970, 1971; Yntema & Torgerson, 1961)."

Summary

In summary, the previous discussion made a case for the use of information utility as a proxy measure for strict information value in DORAC evaluations. An overview of utility theory and the technology of utility measurement was also presented. These results lead to the following conclusions with respect to the use of MAUM in CIEA:

1. Problem structuring is the most critical aspect of utility measurement. This phase of the analysis is also the least amenable to formal prescription. Currently, no universally applicable procedures for utility problem structuring exist.

2. In terms of partial utility assessment, Direct scaling methods have been found to be most useful. The Direct methods are flexible in that they are applicable to objective as well as subjective situations. Direct scaling methods also result in quantitative as opposed to ordinal scale scores.

3. Simple additive utility aggregation functions have been found to be more useful than more complex combination functions. One simple additive aggregation model that holds promise for use in CIEA is Model 3, linear additive with higher order functions. This aggregation model will accommodate a direct treatment of multidimensional utility factors, a situation almost sure to be encountered in DORAC evaluations.

The next section of the report presents a methodological framework for CIEA that incorporates the MAUM procedures described in this section. The CIEA methodology is developed within the context of a traditional cost-effectiveness analysis methodology outlined in Kazanowski (1968). In the CIEA methodology, the effectiveness measure--information value--is derived through a MAUM procedure using information utility as a proxy measure for strict information value. To illustrate the application of the procedure, an example CIEA is presented with the methodological description.

SECTION 2

COST AND INFORMATION EFFECTIVENESS ANALYSIS

Introduction

The objective of a Cost and Information Effectiveness Analysis (CIEA) methodology is to provide the framework for selecting a preferred DORAC alternative in terms of the value of the information provided (e.g., operational readiness, training status, etc.) versus the cost of developing, implementing, and operating the capability. CIEA is intended to serve as a guide to decision-makers in establishing requirements and in developing and evaluating alternative DORAC concepts for any parent materiel system.

As a methodology, CIEA is a member of a set of procedures generally known as cost-effectiveness analysis[1]. The term cost-effectiveness denotes a technique in which alternative systems designed to accomplish a goal are evaluated using measures of cost (usually dollars) and separate measures of effectiveness (e.g., reliability, speed, probability of accomplishing a task, or a weighted index of a number of such factors) (Barish & Kaplan, 1978). Under this approach, cost and effectiveness values for each alternative are determined. The systems are then evaluated on the basis of whether the additional benefits of the more effective alternatives are worth their added costs. The use of cost-effectiveness analysis is common in the evaluation of military materiel and training systems (for example, see TRADOC Pamphlet 11-8 or TRADOC Pamphlet 71-10).

The CIEA methodology described in this section is developed within the framework of a general cost-effectiveness methodology outlined in Kazanowski (1968). The phases of the methodology are listed as follows:

1. Assess needs and constraints
2. Define system objectives
3. Identify operational requirements for the attainment of system objectives
4. Develop alternative system concepts
5. Establish system evaluation criteria
6. Generate systems-versus-criteria array
7. Perform sensitivity analyses
8. Select preferred alternative

---

[1]Technically, CIEA as outlined in this report is a *cost-benefit* analysis because a number of effectiveness measures are condensed into a single composite benefit measure that serves as the basis for the evaluation of alternatives.

In line with the discussion presented in Section 1, a MAUM approach to assessing information worth is integrated into the CIEA methodology. The MAUM process used in CIEA is adapted from similar procedures outlined in Edwards (1976) and Johnson and Huber (1977). The corresponding phases in the MAUM process are listed as follows:

1. Determine the issues to which utility ratings are relevant.
2. Determine the relevant factors on which alternative system concepts are to be assessed.
3. Identify the perspective from which utility is to be assessed.
4. Develop operational measures for each factor.
5. Identify the alternatives to be evaluated.
6. Derive importance weights for each factor.
7. Assess utility of each operational measure.
8. Obtain system effectiveness ratings.
9. Determine partial utility scores for each factor.
10. Aggregate partial utilities to produce global utility score for each alternative.
11. Estimate cost for each alternative.
12. Select a preferred alternative.

Figure 2-1 presents the phases in the MAUM procedure keyed to the phases in CIEA (denoted MAUM-CIEA).

In this report, the emphasis of the methodological discussion is on the conduct of CIEA for fielded materiel/training device systems. Procedures for the conduct of CIEA on unfielded systems represent a subset of the complete process, but are not addressed explicitly in this report. The complete CIEA procedure is presented and discussed in the remainder of this section.

Analytic Methodology

The following subsection presents a discussion of the activities involved in the conduct of a CIEA. To aid in the explication of the methodology, an exemplary analysis (the boxed-in sections) is presented with the narrative description for each phase. Each of the distinct phases of CIEA is now discussed in turn.

Assess Needs and Constraints

Phase one of the CIEA process involves establishing the need for a DORAC and defining general constraints that will serve as a guide for the analysis. The impetus for DORAC development will generally come from a

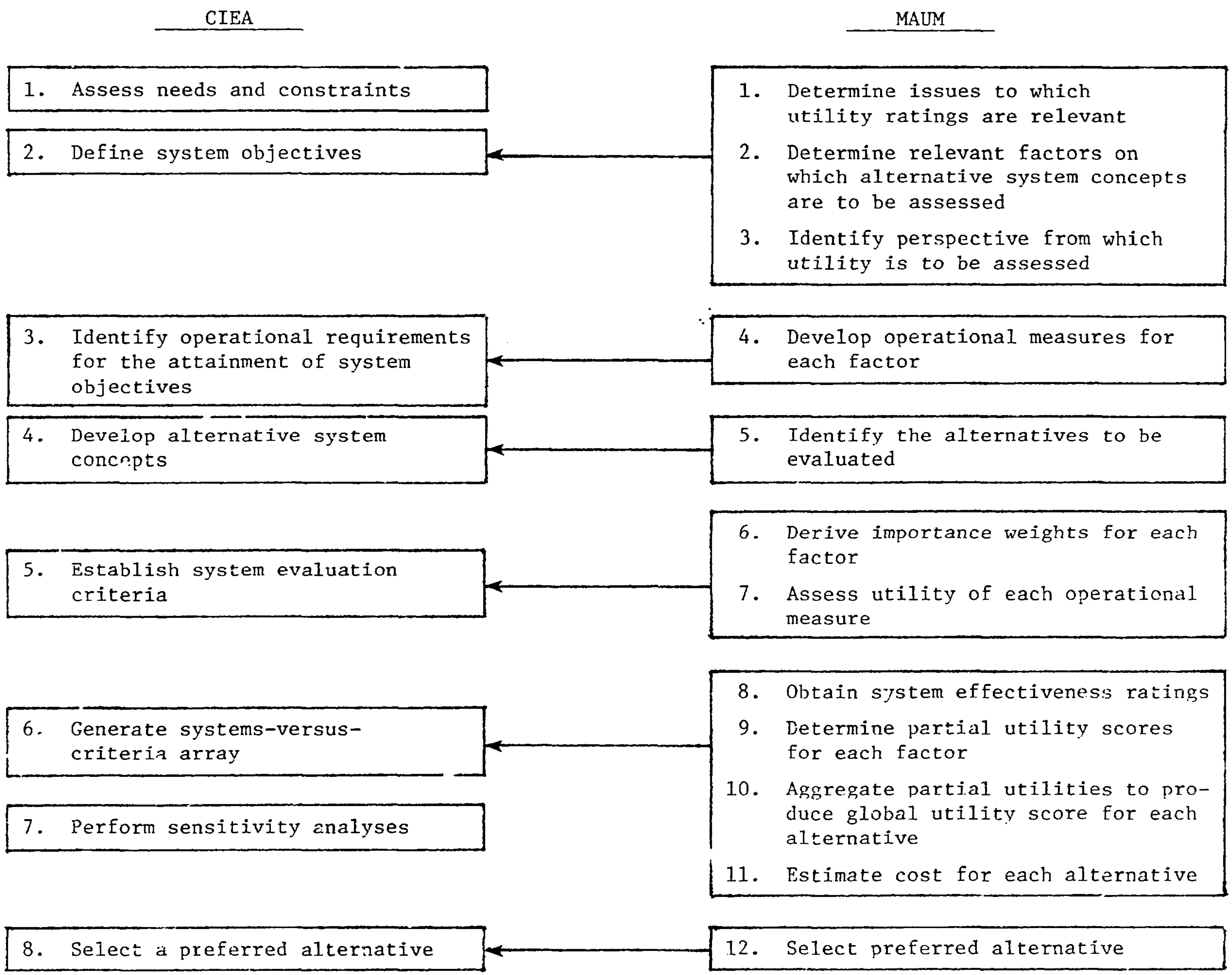

Figure 2-1. The MAUM-CIEA Process

leadership concern that readiness evaluation or reporting for a particular materiel system (e.g., M-16 Rifle, Infantry Fighting Vehicle, etc.) is not being done adequately. Leadership concern may arise from a number of sources. For example, it may result from the fact that Unit Status Reports (USRs) do not agree with Skill Qualification Test (SQT) results, Army Training and Evaluation Program (ARTEP) reports, or results from other individual or unit exercises. In other situations, the judgment that "things are not right" may be based primarily on commanders' subjective opinions. In yet other cases, ammunition and other costs and constraints may limit the frequency with which information on current status can be obtained. Whatever the source or basis, the impetus for the development of a DORAC will be generated by the identification or perception of a possible readiness deficiency that is judged to have significant impact on the Army's fighting ability.

After the need for a DORAC is established, the next aspect of needs/constraints assessment is a formal recognition of need in the form of a problem statement. The problem statement should define, in general terms, the basis for leadership concern and address possible sources of the perceived inadequacies in the readiness evaluation or reporting systems (e.g., no data are collected, the wrong data are collected, the data that are collected are invalid or unreliable, data are not collected often enough, etc.).

CIEA must address a real world situation in which constraints exist. Thus, after recognizing the need for improved readiness reporting, a third aspect of needs/constraints assessment is to identify potential constraints on the development or deployment of such a DORAC. The principal categories of constraints that should be considered include the following (TRADOC Pamphlet 11-8):

1. Economic. The DORAC must be developed and implemented within the context of the economics that apply to the time period being considered.

2. Military. Complex alternatives, no matter how attractive, will usually have to be operated and maintained by Army personnel.

3. Technological. The DORAC should be constrained by the use of demonstrated technology. Alternatives that push the technical state of the art should be given careful review. Unattainable technological objectives can result in the loss of time, money, and effort.

4. Personnel. Personnel resources are scarce even in wartime. Evaluation systems that require significantly more personnel than those they replace should be considered carefully.

5. Time. All changes require time to implement. The CIEA should consider the time frame for the development and deployment of a DORAC. These times should be evaluated with an eye toward the developmental stage of the parent materiel system. For example, if a materiel system is scheduled for replacement beginning in fiscal year (FY) 1983, it makes little sense to develop a DORAC that will not be fully deployed until FY 1982, unless the capability can also be used with the system's replacement.

It is doubtful that all relevant constraints can be identified during the needs/constraints phase of the analysis. However, applicable categories of constraints should be specified. One constraint that should be considered during needs/constraints assessment is system cost. Cost will constrain the kinds of DORACs that are deployed. Hence, benchmark cost guidelines should be provided early in the analysis. Early determination of cost constraints will serve to eliminate excessively costly alternatives early in their developmental cycle.

Define System Objectives

The second phase in CIEA is to refine the problem statement developed in the previous phase into a formal specification of the objectives of the proposed DORAC. Phase 2 is carried out in three separate steps, listed as follows:

1. Determine the issues to which DORAC information is relevant.
2. Specify information worth dimensions.
3. Identify the perspective from which information worth is to be assessed.

Step one involves answering the question: "What purposes are the DORAC data to serve?" This question is addressed by developing a list of information issues (e.g., can our riflemen be used effectively in assault operations? Can our riflemen be expected to engage hostile targets presenting the greatest threat?) from the problem statement. The information issues are similar in purpose to the essential elements of analysis in Cost and Training Effectiveness Analysis (CTEA); they serve to guide the definition of criteria for information worth evaluation.

After a list of information issues has been developed, the second step of phase 2 is to organize the information issues into a set of worth dimensions (WDs), or categories of information use. The WDs constitute the primary value factors for the evaluation of DORAC alternatives. Examples of some potential DORAC worth dimensions are provided as follows:

1. Readiness Evaluation. The determination of whether or not individuals/units are capable of performance at an acceptable level/standard on performances specific to the DORAC information issues.

2. Training Management. The use of training status and performance diagnostic information in determining who, how often, when, and what to train for individual/unit performances related to the specific DORAC.

3. Unit Management. The use of objective job performance information to provide guidance in various unit management activities such as the award of performance incentives, the assignment of personnel to critical unit positions, and so forth.

4. Fighting System Evaluation/Development. The use of evaluation data to provide feedback to branch schools and other concerned agencies on training program content, training materials, training devices, system equipment, support equipment, doctrine, tactics, and so forth.

Progressing from information issues to WDs is likely to lead to the identification of additional information issues. Hence, phases one and two should be performed iteratively until all pertinent information issues and WDs are identified.

In developing WDs it is necessary not to be too expansive. For reasons that will be made apparent later, the number of WDs should not exceed seven. If more than seven WDs are developed, the list should be reviewed and the number of WDs reduced by redefining and combining dimensions. For most applications, the WDs listed above should suffice.

The last step in phase 2 involves identifying the perspective from which information utility is to be assessed. Considering the WDs developed in step 2, organizations having a stake in determining the value of DORAC information are specified. When several organizational perspectives are indicated, persons who can speak for the various stakeholders must be identified and induced to participate in the evaluation.

<u>Identify Operational Requirements for the Attainment of System Objectives</u>

Operational requirements are derivatives of system objectives that specify precisely what measurement capability is required on the part of a DORAC. The objective of phase 3 is to translate the system objectives (i.e., information issues) developed in phase 2 into specific operational requirements for the DORAC. This includes specifying: (1) What performances are to be evaluated, (2) the conditions under which the performances must be demonstrated, and (3) the standards by which performances are evaluated. Each of these aspects of the operational requirements specification is now discussed separately.

Performances. The first step in phase 3 is to state precisely *what* performances are implied in the information issues. These should be stated as a list of relevant performance statements. Examples of acceptable performance statements are listed as follows:

- Climb the telephone pole
- Disassemble an M-16A1 rifle
- Camouflage the helmet
- Add two five-digit numbers

Note that each performance statement includes an *action* verb. The action verb is the key to the performance; it tells exactly what must be done. For example, in the statement "Disassemble an M-16 A1 rifle," the action verb is "disassemble". It is possible to assess directly an individual's ability to perform this action.

Conditions. Each performance statement must include a statement of the conditions under which the performance is to be demonstrated. The condition statement should indicate the following aspects of each performance:

1. What an evaluatee has to work with (e.g., tools, reference materials, etc.).

2. The environmental circumstances under which the performance must be demonstrated (e.g., nighttime, daylight, rain, etc.).

3. What the evaluatee must work on--his starting point (e.g., "given an M-16A1 rifle").

4. Any limitations or special considerations.

It is important to specify all conditions that are relevant to performance assessment using the DORAC. In many cases, performances may be demonstrated under a number of conditions. The statement of conditions pertinent to DORAC evaluation should make clear which conditions are to be considered.

Standards. Finally, each performance statement must specify the standard or criterion by which the performance is evaluated. The standard indicates *how well* or *how quickly* (or both) a performance is to be done. Six categories of standards are usually employed to indicate how well or how quickly a performance must be done. These categories of standards are listed as follows (Swezey & Pearlstein, 1974):

1. Standard Operating Procedure (SOP)--performance must match a specified SOP. The standard specifies when a performance is complete and its proper sequence.

2. Zero error--performance must be completed with 100% accuracy.

3. Minimum acceptable level--performance must meet a specified minimum acceptable level.

4. Subjective quality--performance must match stated qualitative characteristics.

5. Time requirements--performance must be done within a given time period.

6. Production rate--performance must yield a specified output per unit of time.

In most situations involving fielded materiel systems, existing task analysis documentation will provide the information necessary to develop performance statements, conditions, and standards (i.e., performance objectives). Situations may be encountered, however, (e.g., when working with an unfielded materiel system) in which performance objectives will have to be developed by the analyst. As in CTEA, performances, conditions, and standards for unfielded systems can be developed: (1) from a knowledge of antecedent systems, (2) on the basis of available system development documentation [e.g., LSAR (Logistics Support Analysis Record) documentation], or (3) from judgments rendered by subject matter experts.

As noted previously, performance objectives (i.e., tasks, conditions, and standards) are derivatives of the information issues developed in phase 2. The performance objectives constitute an operational definition of precisely what is implied in each of the information issues. As such, the performance objectives represent the operational measures upon which the evaluation of each DORAC alternative is based. The application of each component of the performance objectives are addressed later in the report at appropriate places in the discussion. For clarification purposes, Figure 2-2 depicts the relationship between the problem statement, WDs, information issues, and performance statements, conditions, and standards.

Develop Alternative System Concepts

Phase 4 is concerned with characterizing alternative DORAC concepts for meeting the operational requirements developed in Step 3. In this step, information concerning: (1) what is required to be measured (i.e.,

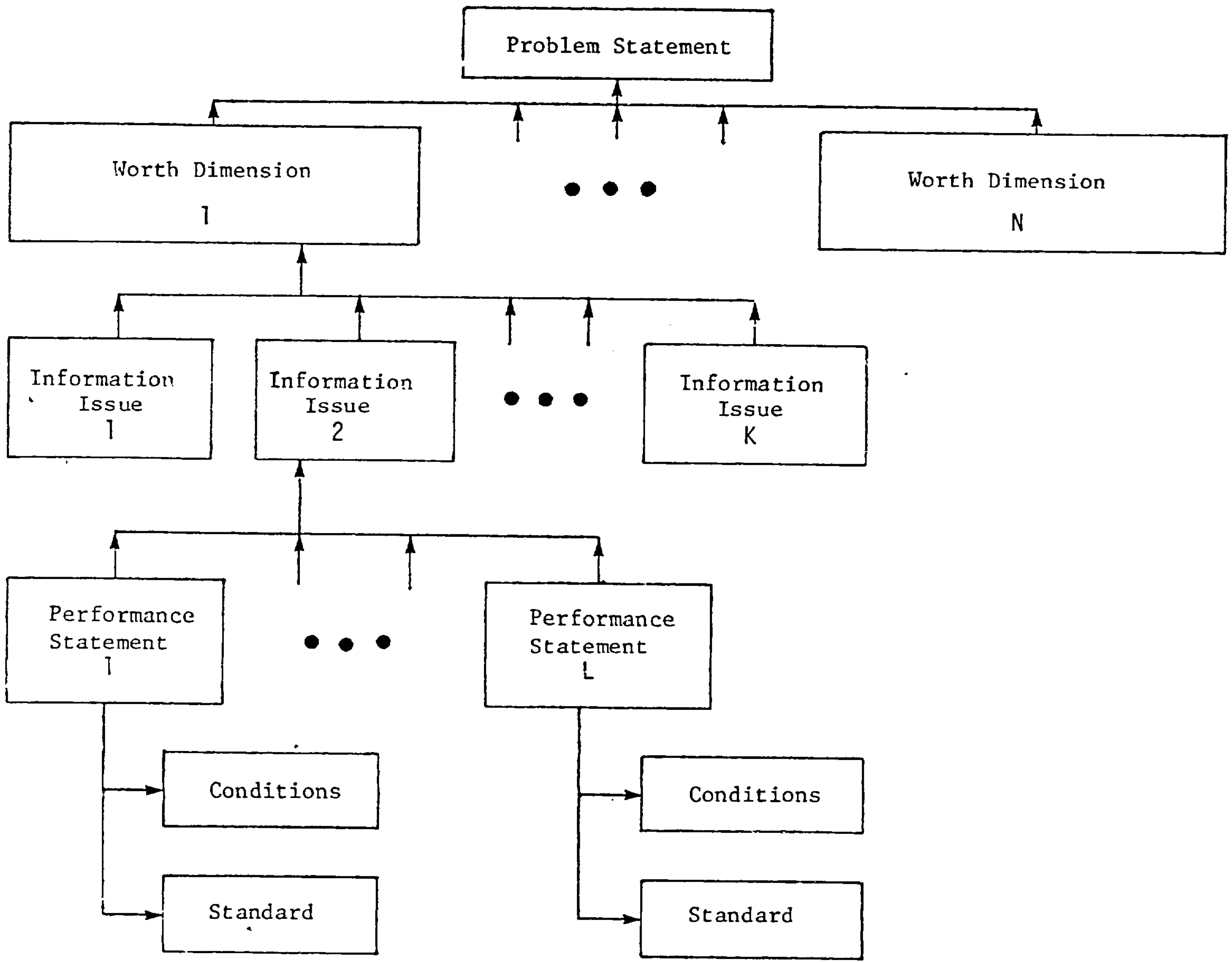

Figure 2-2. The Relationship Between the Problem Statement, Worth Dimensions, Information Issues, and Performance Objectives

performance statements), (2) the conditions under which performances are to be done, and (3) performance standards are used to specify operational measurement capabilities. Differences in operational measurement capabilities define DORAC alternatives.

In considering the operational characteristics of DORAC alternatives, the first differentiating variable is performance assessment capability; that is, whether or not an alternative provides a vehicle for assessing a particular performance. To document assessment capabilities, DORAC alternatives are scored using a Performance by Alternative matrix. Entries in this matrix are either "X" or "O", indicating that an alternative either does or does not provide a vehicle for assessing a given performance. Figure 2-3 presents a hypothetical Performance by Alternative matrix.

The second differentiating variable for DORAC alternatives involves performance conditions. Working from the performances that can be assessed using one or more of the DORAC alternatives, the relevant conditions for each are listed. Each DORAC alternative is then further characterized according to whether or not it provides a vehicle for assessing a given performance under each relevant condition. Again, an "X" or "O" scoring scheme is used to characterize each alternative. The conditions data are recorded in the Performance by Condition matrix. A separate Performance by Condition matrix is produced for each DORAC alternative. Figure 2-4 presents an exemplary Performance by Condition matrix.

The previous incidence matrices (Figures 2-3 and 2-4) characterize DORAC alternatives according to their assessment potential; that is, by whether or not it is possible to assess specific performances under given conditions using a particular alternative. Along with the incidence data, it is also necessary to specify *how* performance assessment is done. Accordingly, each cell in the Performance by Alternative matrix containing an "X" (i.e., assessment is possible) is next characterized according to how performance assessment is accomplished. For example, potential methods of performance measurement include the following:

1. Observer verification,
2. Performance checklist,
3. Numerical rating scale,
4. Machine recording with observer summary report,
5. Machine recording with hardcopy trace and summary report.

Each assessment method is rated according to the judged *dependability* of the data it will provide. Dependability ratings are made using the following 5-point scale, presented as follows:

Alternative

Performance		A	B	C	● ● ●	Z
	1	X	X	X		O
	2	X	O	O		X
	3	O	X	X		X
	● ● ●					
	i	X	X	X		X
	● ● ●					
	N	O	O	X		X

Figure 2-3. Hypothetical Performance by Alternative Matrix

Alternative 0:

Condition

Performance

	A	B	C	• • •	Z
1	X		X		
2	X				
3			X		X
• • •					
i	X	X			X
• • •					
N	X	X	X		X

Figure 2-4. Hypothetical Performance by Condition Matrix

1. Very low
2. Low
3. Moderate
4. High
5. Very high

Factors that should be considered in assigning dependability ratings include:

1. The judged validity of the resulting data. Is the recorded score likely to be a valid characterization of true evaluatee performance?

2. The inherent stability (i.e., reliability) of the method.

3. The judged precision of the method. Does the assessment methodology provide data with sufficient precision to establish whether or not the performance standard is met?

4. The amount of human intervention in the measurement process.

Data dependability is directly proportional to the first three factors and inversely related to the third.

The last step in Phase 4 involves developing an evaluation scenario for each DORAC alternative. This scenario outlines characteristics of the evaluation situation such as: (1) whether or not all performances are to be assessed during each evaluation (this will impact upon the time required to conduct an evaluation), (2) the number of planned evaluations per year, (3) the anticipated number and type of evaluators per evaluation position, (4) whether or not the collection of evaluation data is part of regularly scheduled training, and so forth. When combined with the operational characteristics of each proposed alternative, the evaluation scenarios provide the basis for generating cost estimates for DORAC alternatives. The position requirements portion (i.e., steps 1 through 3) of the DORAC Costing Guidelines presented in Appendix A should prove useful in developing evaluation scenarios.

At this point in the discussion, an illustration CIEA is introduced. The example exercise is structured around the conduct of an analysis for a hypothetical set of training device alternatives for the M-16A1 rifle plus the baseline field "record fire" procedures. It should be noted that a *baseline* case consisting of the *current* evaluation procedure will often be used in CIEA. For purposes of analysis, a needs/constraints assessment (phase 1) is assumed; it is further assumed that information issues have been specified and have led to the development of two WDs, listed as follows:

1. Readiness Evaluation
2. Training Management

An analysis of DORAC objectives has resulted in the following performance statements:

1. Load/unload rifle magazine
2. Reduce stoppage and clear rifle
3. Zero rifle
4. Engage stationary targets under field-fire (FF) conditions
5. Engage moving targets under FF conditions
6. Engage moving targets from defensive position
7. Engage moving targets in assault mode

The performance statements categorized by WDs are listed as follows:

Readiness Evaluation

1. Engage stationary targets under FF conditions.
2. Engage moving targets under FF conditions.
3. Engage moving targets from defensive position.
4. Engage moving targets in assault mode.

Training Management

1. Load/unload rifle magazine.
2. Reduce stoppage and clear rifle.
3. Zero rifle.
4. Engage stationary targets under FF conditions.
5. Engage moving targets under FF conditions.

Performance conditions and standards, other than those implied in the performance statement, are not explicitly considered in the example exercise.

Phase 4, Develop Alternative System Concepts, is used to characterize three hypothetical alternatives, denoted B (baseline), X, and Y. Figure 2-5 presents the Performance by Alternative matrix for the three hypothetical alternatives.

			Alternative		
			B	X	Y
	1.	Load/unload	X	X	X
	2.	Reduce stoppage	X	X	X
	3.	Zero	X	X	X
Performance	4.	Stationary--FF	X	X	X
	5.	Moving-FF	O	O	X
	6.	Moving-defense	O	O	X
	7.	Moving-assault	O	O	X

Figure 2-5. Performance by Alternative Matrix for CIEA Example Exercise

Measurement method and data dependability evaluations for the three alternatives are presented in Figure 2-6.

			Alternative		
			B	X	Y
	1.	Load/unload	Observer verification(OV)/ Moderate (Mod)	OV/ Mod	OV/ Mod
	2.	Reduce stoppage	OV/ Mod	OV/ Mod	OV/ Mod
	3.	Zero	OV/ Mod	Machine scored/ High	Machine scored/ High
Performance	4.	Stationary--FF	OV/ Mod	Machine scored/ Mod	Machine scored/ High
	5.	Moving--FF	---	---	Machine scored/ High
	6.	Moving-defense	---	---	Machine scored/ High
	7.	Moving-assault	---	---	Machine scored/ High

Figure 2-6. Measurement Method/Data Dependability Evaluations for CIEA Example Exercise

The final product of phase 4 is the evaluation scenario. It was specified that each performance be evaluated during each evaluation period; furthermore, each alternative is to be employed with the following frequency:

	Alternative	Frequency
1.	B (Baseline)	Once per year
2.	X	Twice per year
3.	Y	Quarterly (four times per year)

Establish System Evaluation Criteria

The next phase in CIEA involves rating the system evaluation criteria--WDs and OMs--in terms of their importance and utility, respectively. First, the WDs are rated according to their importance to overall DORAC value. In step two, the OMs nested within WDs, are rated according to their utility for decision-making. The rating techniques employed are adapted from similar procedures presented in Churchman, Ackorf, and Arnoff (1957), Edwards (1976), and Barish and Kaplan (1978).

Worth Dimensions. The first step in establishing system evaluation criteria involves determining the importance of each of the WDs to overall DORAC value. Importance weights for the WDs are obtained using the series of substeps presented below:

1. Rank the WDs in order of importance. The ranking can be performed either by an individual, or by representatives of various stakeholders acting as a group. The group process can be formal, as in Delphi or the Nominal Group Technique, or informal. If a group procedure is employed, the final rankings should reflect a consensus of opinion.

2. Rate the WDs on importance, preserving ratios.

    a. Assign the least important WD a rating of 10.

    b. Consider the next-least-important WD. How much more important is it than the least important? Assign it a number that reflects that ratio. For example, if the second-least-important WD is judged to be four times as important as the first, it is assigned a score of 40. Continue up the list of WDs, checking each set of ratios as each new judgment is made.

c. Sum the resulting importance scores, divide each by the sum, and multiply by 100.

3. Reevaluate importance ratings to improve reliability.

   a. Compare the first (most important) WD with the remaining ones put together. Is it more important, equally important, or less important than all the others put together?

   b. If the first WD is <u>more</u> important than all of the others put together, see if it's importance rating is greater than the sum of the importance ratings for all of the other WDs. If not, change the importance rating of the first WD so that it is greater than the sum of the others.

   c. If the first WD is of <u>equal</u> importance to all the others put together, see if its importance rating is equal to the sum of the importance ratings of all the other WDs. If it is not, change the importance rating of the first WD so that it is equal to the sum of the others.

   d. If the first WD is <u>less</u> important than all the others put together, see if its importance rating is less than the sum of the importance ratings of all of the other WDs. If it is not, change the importance rating of the first WD so that it is less than the sum of the others.

   e. If the first WD was considered <u>more</u> important or <u>equally</u> important than all the others put together, apply the above procedure to the second-most-important WD on the list. Is it more important, equally important, or less important than all the other farther down the list put together? Then, proceed as in (b), (c), and (d) above, applying the revision procedure to the second WD instead of the first.

   f. If the first WD was considered <u>less</u> important than all the others put together, compare the first WD with all the remaining ones put together, except the lowest rated one. Is the first WD more important, equally important, or less important than all of the others farther down the list except the lowest one put together? Then proceed as in (b), (c), and (d) above. If (b) or (c) are applicable, proceed to (e) after applying (b) or (c). If (d) is applicable, proceed as in this paragraph (f) again, comparing the first WD with all the remaining ones put together except the lowest two. As long as (d) is applicable, the procedures of this paragraph (f) are repeated until the first WD is compared with the second and third WDs put together. Then, even if (d) is still applicable, proceed to (e).

g. Continue the above procedure until the third-from-the-lowest WD has been compared with the two lowest WDs on the list.

4. Repeat action (2-c) on the revised importance ratings produced in substep (3) except do not multiply by 100. The results of this substep (4) are the importance weights for each WD.

A sample Rating Development Sheet (RDS) is shown as Figure 2-7. On the RDS, space is provided for no more than seven WDs. When more than seven WDs are included, the rating procedure becomes cumbersome. If more than seven initial WDs are developed, the decision-making individual or group should review the list of WDs and reduce their number to seven or fewer by eliminating the least important dimensions or by combining dimensions. If it is not possible to reduce the list of WDs to seven or fewer, a final course of action is to omit substep (3), and retain the initial ratings from substep (2). However, stopping with substep (2) can result in importance weights that are less reliable than weights obtained through the exercise of the entire procedure.

Operational Measures. The second step in establishing system evaluation criteria involves determining the utility, or *value for decision-making*, of the OMs nested within each of the WDs. Utility scores for OMs are determined using the following series of substeps:

1. List the OMs in descending order of utility (i.e., value for decision making) to the WD being considered.

2. If there are seven or fewer OMs nested within a WD, obtain utility scores for each following the procedure outlined in the previous section (i.e., for weighting WDs).

3. If a WD is characterized by more than seven OMs, obtain utility scores as follows:

a. Select one OM at random.

Worth Dimension	Initial Rating (from sub-step 2)	Revised Ratings (Iterations of substep 3)										Importance Weight (from sub-step 4)
		1	2	3	4	5	6	7	8	9	10	
1. WD_1												
2. WD_2												
3. WD_3												
4. WD_4												
5. WD_5												
6. WD_6												
7. WD_7												

Figure 2-7. Sample Rating Development Sheet

b. Randomly assign each of the <u>remaining</u> OMs to groups of approximately equal size, with no more than five OMs to a group.

c. Add the selected OM (a) to each group and assign it a rating of 100. This OM will serve to link each of the groups for a later recombination of the ratings (action 3-f).

d. Rank each of the OMs in each group in order of descending utility. Then, assign ratings to them following substeps 2 and 3 of the WD weighting procedure outlined previously. In each group, keep the rating of the selected OM (a) at 100.

e. For each group, follow the procedure described in substep 3 of the WD weighting procedure. Do not change the rating of 100 assigned to the selected OM (a) and do not adjust the ratings (substep 4).

f. Make a combined listing of all the OMs in order of decreasing utility. Compare this listing with the initial rankings from substep 1. Note any difference in rankings. If the initial list (1) is judged correct, repeat actions (a) through (e) to adjust the affected groups and reconcile the evaluations.

g. Adjust the utility ratings as described previously (action 2-c, p. 2-17). Do not multiply the results by 100.

4. Repeat substeps 1 through 3 for each WD.

The results of CIEA phase 5 are a set of importance weights for WDs and a set of utility scores for OMs nested within WDs. These values are used later in the analysis to produce partial utility scores for each WD, and to combine partial utility scores to produce global utility scores for DORAC alternatives.

To illustrate phase 5, the example CIEA is continued in the following paragraphs. The first step involves rating the WDs according to their importance to overall DORAC value. Accordingly, Readiness Evaluation (RE) and Training Management (TM) are ranked one and two, respectively. Furthermore, RE is judged to be four times as important as TM. The results of step one are summarized as follows:

Worth Dimension	Rank	Initial Importance Rating	Scaled Importance Rating
Readiness Evaluation	1	40	.80
Training Management	2	10	.20

Step two of phase 5 is concerned with establishing the utility for decision-making of each OM nested within each WD. This step is conducted in two substeps. The first substep is to obtain a set of initial scaled utility ratings by applying the rank-rate procedure outlined above. Results from substep one are presented as follows:

Worth Dimension	Rank	Initial Utility Rating	Initial Scaled Utility Rating
Readiness Evaluation			
1. Moving-assault	1	100	43
2. Moving-defense	2	75	32
3. Moving-FF	3	50	21
4. Stationary-FF	4	10	4
Training Management			
1. Moving-FF	1	200	54
2. Stationary-FF	2	100	27
3. Zero rifle	3	50	14
4. Reduce stoppage	4	10	3
5. Load/unload	5	10	3

Following the development of an initial set of scaled utility ratings for the OMs, substep two of step two is initiated. The objective of substep two is to refine the output of substep one into utility scores. Considering, for example, the initial scaled utility ratings for the OMs nested under Readiness Evaluation (see the table below), *moving-assault* is judged equal in utility to the remainder of the OMs. Thus, the utility rating for moving assault is changed from 43 to 57, the sum of the ratings for the other OMs. Proceeding down the list, moving-defense is judged to have more utility than the remaining two OMs. Its rating of 32 is greater than 21 + 4 = 25; hence, the utility rating for moving-defense does not change. The final action is to rescale the final set of revised utility ratings to obtain utility scores (.50, .28, .18, .04). A similar process is used to obtain the utility scores for Training Management. The complete set of results from substep two for the example CIEA is given as follows:

Worth Dimension:	Initial Scaled Ratings	1	2	3	4	5	Utility Scores
Readiness Evaluation							
1. Moving-assault	43	57					.50
2. Moving-defense	32	32					.28
3. Moving-FF	21	21					.18
4. Stationary-FF	4	4					.04
Training Management							
1. Moving-FF	54	94	94	120			.67
2. Stationary-FF	27	27	40	40			.22
3. Zero rifle	14	14	14	14			.08
4. Reduce stoppage	3	3	3	3			.02
5. Load/unload	3	3	3	3			.02

Generate Systems-Versus-Criteria Array

Following the generation of importance weights for WDs and utility scores for OMs nested within WDs, the next phase in MAUM-CIEA involves generating the systems-versus-criteria array. In CIEA, the systems-versus-criteria array is a matrix that presents each alternative and its associated evaluation criteria--Information Utility, Relative Information Utility, Information Cost, Relative Information Cost, and Relative Information Worth.

Developing the systems-versus-criteria array is a five step procedure, with the steps listed as follows:

1. Obtain System Effectiveness Ratings
2. Determine Partial Utilities for WDs
3. Aggregate Partial Utility Scores
4. Estimate Costs of Alternatives
5. Determine Relative Information Worth

Each of these steps is discussed in the following paragraphs.

Obtain Systems Effectiveness Ratings. Within the context of CIEA, system effectiveness is defined as the degree to which an alternative provides timely, quality information on the performances relevant to the DORAC evaluation. Specifying system effectiveness is carried out in three substeps: First, information *quality* ratings are obtained for each alternative on each OM. Next, each DORAC alternative is evaluated with respect to the utility of the *frequency* with which performance data are provided. Finally, the quality ratings and frequency utility scores are combined to produce a single measure of the *effectiveness* of each alternative on each OM nested within each WD. The next series of paragraphs describe the effectiveness rating procedure in additional detail.

Information quality is defined as the extent to which an alternative provides *trustworthy* information relevant to a particular OM. Also considered as part of information quality is the amount of information provided; that is, the number of relevant performance conditions that are addressed by the alternative. In CIEA, information quality ratings are obtained using the procedure outlined as follows:

1. Order the DORAC alternatives from "best" to "worst" according to the degree to which each alternative is capable of providing quality information relevant to the OM under consideration. Factors that should be considered in making quality judgments include:

   a. Amount of information. The number of relevant performance conditions that are addressed.

   b. Dependability. The judged trustworthiness of the data. This is obtained from the dependability ratings made previously (e.g., Figure 2-6).

   Ties are permitted. If one or more of the alternatives are judged equivalent in terms of the quality of the information they provide, assign them the same rank.

2. Numerically position the best and worst alternatives on a 0-to-100 scale. Use the following benchmark ratings as a guide:

   0 - The alternative provides no data relevant to the OM under consideration.

   25 - Marginal. The alternative provides partial data on the OM and the data are likely to be undependable (e.g., the recording/scoring method is poor resulting in low validity or low reliability).

50 - Adequate. The alternative provides the required data but some dependability problems are apparent. For example, the most appropriate recording/scoring method is not used or the data are likely to have only moderate reliability.

75 - Good. The alternative provides required data in an acceptable manner. Recording methods are acceptable; reliability is likely to be reasonably high.

100 - Excellent. The alternative is the <u>best possible</u>, given the current technical state of the art. Recording methods are automated and precise; reliability is likely to be very high.

3. Position the remaining alternatives between the best and worst cases on the 0-to-100 scale. Again, refer to the benchmark rating points presented in (2) as a guide.

Continuing with the example CIEA, quality ratings for the three hypothetical DORAC alternatives are presented in Table 2-1.

Table 2-1. Quality Ratings for Hypothetical DORAC Alternatives

		Alternative B	Alternative X	Alternative Y
Operational Measure	<u>Readiness Evaluation</u>			
	Moving-Assault	0	0	25
	Moving-Defense	0	0	50
	Moving-FF	0	0	50
	Stationary-FF	50	50	75
	<u>Training Management</u>			
	Movin[illegible]	0	0	50
	Stationary--FF	25	50	40
	Zero rifle	25	50	50
	Stoppage	25	25	25
	Load/Unload	25	25	25

After information quality ratings have been assigned for each alternative on each OM, the next substep in obtaining effectiveness ratings is to determine the utility of the evaluation frequency associated with each alternative. Frequency utility ratings are obtained by applying the following sequence of actions:

1. Consider the frequency of the information provided by each alternative (e.g., quarterly, twice a year, yearly, etc.). Now, specifically considering the highest and lowest frequencies, rate the usefulness of receiving DORAC generated information with the frequencies indicated. Use a 0-to-100 scale in assigning these ratings.

2. Next, consider the remaining intermediate frequencies. Position the remaining frequencies between the extreme values (i.e., ratings for the highest and lowest frequencies) on the 0-to-100 scale.

3. Connect the scaled points with a line. If the utility rating for the lowest frequency is not zero, connect the point associated with the lowest frequency with the zero point on the frequency-utility axes. The result is the evaluation frequency utility curve for the alternatives under consideration.

As an example of the frequency utility rating process, consider the evaluation frequencies associated with the DORAC alternatives in the example CIEA. The frequencies for alternatives B, X, and Y are 1, 2, and 4 times per year, respectively. One (the lowest value) is assigned a utility rating of 50; four (the highest value) is assigned a score of 75. The only intermediate value, two, is assigned a rating of 70. The resulting utility curve for evaluation frequency is shown in Figure 2-8.

The final substep in the process of obtaining effectiveness ratings for alternatives is to combine the quality and frequency ratings into a single measure of system effectiveness. Quality and frequency are combined using the formula:

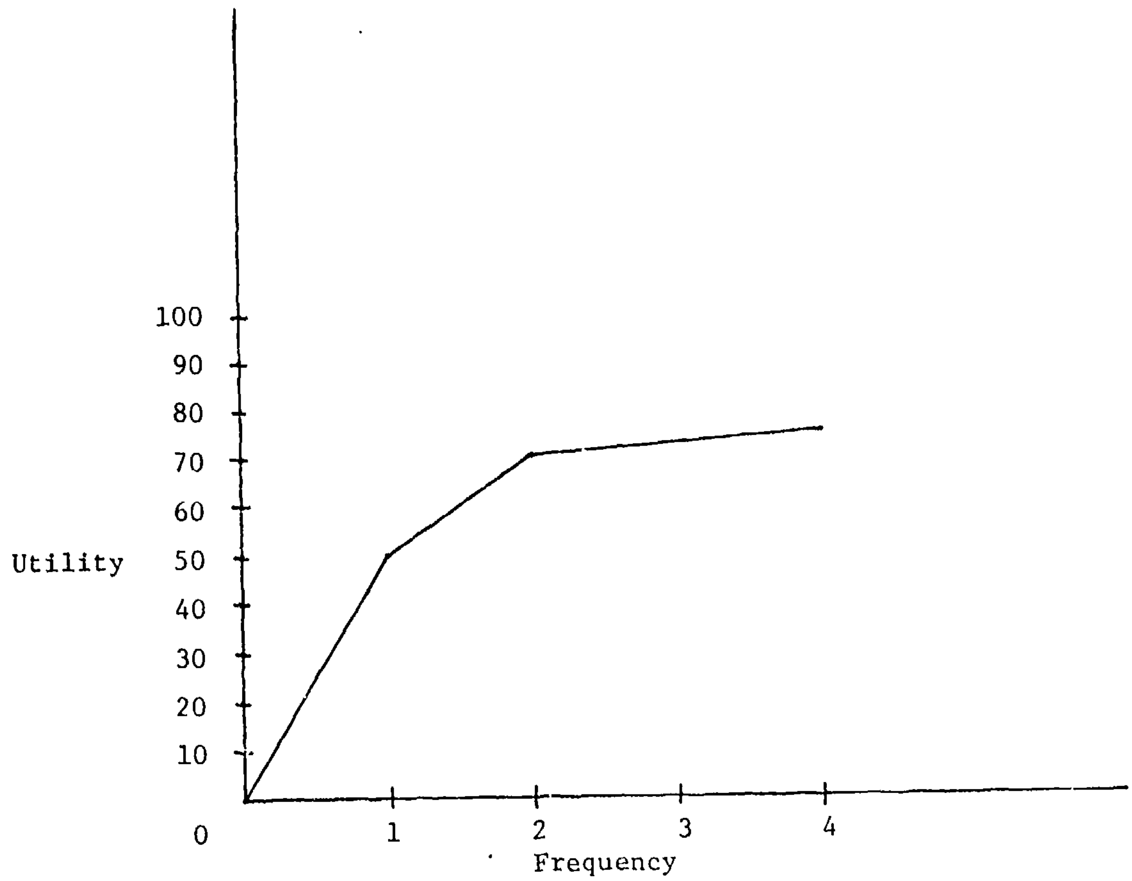

Figure 2-8. Utility Curve for Evaluation Frequency

$$\varepsilon_{ijk} = Q_{ijk} \cdot F_i/100, \tag{2-2}$$

where ε_{ijk} represents the effectiveness of the i^{th} DORAC alternative on the k^{th} OM nested within the j^{th} WD,

Q_{ijk} is the quality of data provided by the i^{th} alternative on the k^{th} OM nested within the j^{th} WD,

and F_i is the frequency utility rating associated with the i^{th} DORAC alternative.

The result of the effectiveness evaluation procedure is a systems effectiveness matrix for each WD. Figure 2-9 presents an example of a hypothetical systems effectiveness matrix for WD_j. The cell entries, ε_{ki}, represent the judged effectiveness scores for DORAC alternatives on the j^{th} WD. If desired, effectiveness scores for all alternatives can be combined across WDs to produce a systems effectiveness array. For example, Table 2-2 presents the systems effectiveness array for the example CIEA.

Table 2-2. Systems-Effectiveness Array for Hypothetical DORAC Alternatives

		Alternative		
		B	X	Y
Operational Measure	Readiness Evaluation			
	Moving-Assault	0	0	18.75
	Moving-Defense	0	0	37.50
	Moving-FF	0	0	37.50
	Stationary-FF	25.00	35.00	56.25
	Training Management			
	Moving-FF	0	0	37.50
	Stationary-FF	12.50	35.00	30.00
	Zero	12.50	35.00	37.50
	Stoppage	12.50	17.50	18.75
	Load/Unload	12.50	17.50	18.75

Worth Dimension j	Alternative				
	1	2	3	• • •	i
OM_1	ε_{11}	ε_{12}	ε_{13}		ε_{1i}
OM_2	ε_{21}	ε_{22}	ε_{23}		ε_{2i}
OM_3	ε_{31}	ε_{32}	ε_{33}		ε_{3i}
• • •					
OM_k	ε_{k1}	ε_{k2}	ε_{k3}		ε_{ki}

Figure 2-9. Hypothetical Systems Effectiveness Matrix

Determine Partial Utilities for WDs. Following the generation of the systems effectiveness array, the next step in the analysis is to produce partial utility scores for each alternative on each WD. The general form of the equation for aggregating effectiveness ratings across OMs is given in (2-3):

$$f[u(x_{ij})] = \sum_k u_{jk} \, \varepsilon_{ijk} \qquad (2\text{-}3)$$

where $f[\cdot]$ is the partial utility score of the i^{th} alternative on the j^{th} WD,

u_{jk} is the utility score of the k^{th} OM nested in the j^{th} WD,

and ε_{ijk} is the effectiveness rating of the i^{th} DORAC alternative on the k^{th} OM nested within the j^{th} WD.

Again continuing the CIEA example, (2-3) is used to combine the utility scores for OMs into partial utility scores for WDs. For example, the partial utility score of alternative B (baseline) for Readiness Evaluation would be derived as:

$$u = (.50)(0) + (.28)(0) + (.18)(0) + (.04)(25) = 1.0 \qquad (2\text{-}4)$$

The complete set of partial utility scores for the hypothetical DORAC alternatives is presented in the table below.

	Alternative		
Worth Dimension	B	X	Y
Readiness Evaluation	1.0	1.4	28.9
Training Management	4.3	11.2	35.5

Aggregate Partial Utility Scores. The next step in generating the systems-versus-criteria array is to aggregate partial information utility scores across WDs to produce a global information utility score for each alternative. In accord with the discussion presented in Section 1, an additive utility aggregation model is recommended. An equation for producing global information utility scores from partial information utilities is presented as follows:

$$IU_i = \sum_j W_j f[u(x_{ij})], \quad (2\text{-}5)$$

where IU_i represents the global information utility score for the i^{th} DORAC alternative,

W_j is the importance weight of the j^{th} WD,

and $f[\cdot]$ is the partial information utility score of the i^{th} alternative on the j^{th} WD.

Following (2-5), the information utility scores for the three hypothetical DORAC alternatives are obtained as follows:

Alternative	IU_i
B	1.66
X	3.36
Y	30.22

As an example of the computation of IU, the score for the baseline alternative (B) is obtained as follows:

$$IU_B = (.80)(1.0) + (.20)(4.3) = 1.66 \quad (2\text{-}6)$$

Estimate Costs of Alternatives. The next-to-the-last step in defining the systems-versus-criteria array involves estimating the cost of each DORAC alternative. Although costing is formally introduced at this point in the discussion, the cost analysis actually may be initiated any time after the alternatives have been specified (i.e., following phase 4, Develop Alternative System Concepts). The cost analysis should, in fact, be initiated as early as possible since this aspect of CIEA will usually prove to be the most time-consuming component.

The objective of the cost analysis substep is to produce a cost estimate for each alternative, denoted C_i. To assist in the development of cost estimates for alternatives, Appendix A presents a structured DORAC Costing Guideline. The guideline leads an analyst through the steps of a DORAC cost analysis beginning with a determination of the anticipated facility load and ending with a total estimated cost for each alternative. It should be noted that the cost estimates provided by the costing guideline consider only those design, development (e.g., testing to validate measures and establish standards), and administration (e.g., testing, data processing) costs which would be incurred over and above those associated with design, development, and use of the devices for training.

Determine Relative Information Worth. Within the standard context of cost-effectiveness analysis (e.g., Barish & Kaplan, 1978; TRADOC Pamphlet 11-8), Relative Information Worth (RIW) is obtained by determining Relative Information Utility (RIU) and Relative Information Cost (RIC) for each alternative and then combining them. RIU for alternative i (RIU_i) is determined by dividing the information utility measure for alternative i (IU_i) by that of another, usually the baseline alternative (i.e., the minimal or least costly DORAC alternative):

$$RIU_i = IU_i/IU_b. \tag{2-7}$$

Relative Information Cost for alternative i (RIC_i) is determined in a similar fashion:

$$RIC_i = C_i/C_b, \tag{2-8}$$

where C_i is the estimated cost of the i^{th} alternative and C_b is the estimated cost of the baseline alternative.

RIU and RIC are useful in themselves, but have limited value in the context of an analysis similar to CIEA where the objective is to identify the most cost-effective alternative. In order to determine a preferred capability, RIU and RIC are integrated into a measure of RIW for each alternative. In CIEA, the RIW of the i^{th} alternative (RIW_i) is defined as follows:

$$RIW_i = RIU_i/RIC_i = \frac{IU_i/IU_b}{C_i/C_b}. \tag{2-9}$$

An RIW score greater than one indicates that alternative i is preferred to the baseline system. In effect, what is done is to normalize system cost and information utility relationships with the baseline alternative assigned a unit value.

To illustrate the development of RIW, again consider the three hypothetical DORAC alternatives. Assume that cost estimates for the three alternatives have been determined as follows:

Alternative	Estimated Cost--C_i
B	$ 175,000
X	525,000
Y	945,000

Applying (2-7) and (2-8), the following RIC and RIU scores are obtained:

Alternative	RIC	RIU
B	1.00	1.00
X	3.00	2.10
Y	5.40	18.20

Finally, using (2-9), RIW for each of the hypothetical DORAC alternatives is determined:

Alternative	RIW
B	1.00
X	0.70
Y	3.37

Based on RIW alone, the results indicate that alternative Y is the most cost and information effective capability.

As a final step in phase 6, all of the system evaluation criteria can be assembled into a formal systems-versus-criteria array. In this array, the rows represent DORAC alternatives and the columns represent the various system evaluation criteria (e.g., IU, RIU, IC, RIC, and RIW). The entries in the matrix are the appropriate evaluation criteria classified by alternative. As an example, the following table is the formal systems-versus-criteria array for the set of hypothetical DORAC alternatives.

	Evaluation Criterion				
Alternative	IU	RIU	IC	RIC	RIW
B	1.66	1.00	175,000	1.00	1.00
X	3.36	2.10	525,000	3.00	0.70
Y	30.22	18.20	945,000	5.00	3.37

Perform Sensitivity Analyses

It is likely that many of the values used in CIEA to specify IU or system cost will be based upon assumption, expert opinion, or incomplete data and thus be of unknown validity. Sensitivity analysis refers to an investigation of the effects on system evaluation criteria of estimated parameters taking on values different from those used in the analysis (Shannon, 1975). Such analyses usually consist of systematically varying the values of selected decision variables over a range of interest and observing the effects of these changes on system evaluation criteria.

Sensitivity analysis can indicate the robustness of the results of a CIEA. It is desirable to determine how far off certain parameter estimates can be without changing the basic conclusions of the analysis. If the results are insensitive to a fairly wide range of changes in selected parameters, then excessive concern need not be given to the accuracy of these parameters. On the other hand, if the results prove to be highly dependent on the values of certain parameters, it may be prudent to expend additional resources and obtain more precise estimates for the parameters in question.

In CIEA, one candidate for sensitivity investigation is the effect of estimated system cost on the selection of a preferred alternative. The first step in the conduct of a cost sensitivity analysis is to establish the cost range over which the top-rated alternative is preferred; that is, to determine the cost estimate that would make the top-rated alternative no longer preferred. Then, determine a pessimistic (i.e., highest) cost estimate for the preferred alternative. The pessimistic cost estimate can be obtained by reverting to the cost determination portion of the CIEA and using pessimistic instead of expected values in the cost analysis. Next, compare the pessimistic cost estimate with the upper bound of the cost range where the top-rated alternative is preferred. If the pessimistic cost estimate is below the upper cost bound, the top-rated alternative is still preferred. However, if the pessimistic cost estimate is above the upper cost bound, then the highest rated DORAC alternatives should be examined in greater detail before selecting a preferred capability.

> To illustrate the cost sensitivity analysis procedure outlined above, consider the top-rated hypothetical DORAC alternative. Alternative Y has an RIU score of 18.20, an RIC score 5.40, and a resulting RIW score of 3.37. The second rated alternative, baseline (B), has RIW, RIU, and RIC scores of 1.00. Accordingly, the RIC score for alternative Y would have to exceed 18.20 for Y's RIW score to be less than that of alternative B. In other words, alternative Y's estimated cost would have to exceed $175,000 X 18.20 = $3,185,000 before the results of the analysis should be reconsidered.

Situations may also arise in which importance ratings for WDs, utility scores for OMs, or effectiveness ratings for DORAC alternatives are at issue; that is, varied opinions concerning the values of certain of these parameters are apparent. An approach to resolving these differences is to conduct a sensitivity investigation generally known as parametric analysis. In a parametric analysis, the values of the parameters in question are systematically varied over a range of interest (i.e., the range dictated by different value structures). The effects of these variations on system selection criteria are observed. If the parametric analysis indicates that different parameter values (worth values) reflecting different points of view result in changes in system rankings, then a decision must be made concerning which value structure is to dominate.

Select Preferred Alternative

The last phase in CIEA is to select a preferred DORAC alternative from among those under consideration. In many situations, the decision rule is simple: maximize RIW, the index of Relative Information Worth. This choice should be made after reviewing the results of appropriate sensitivity analyses.

If the RIW scores for a number of alternatives are virtually identical, the appropriate course of action may be to conclude that there is no difference among the top-rated alternatives. In such situations, the selection of a preferred DORAC alternative may have to be made on the basis of additional analyses or on the basis of considered military judgment.

As a final note, a special selection case arises when one of the evaluation criteria is subject to a constraint (e.g., a budget constraint exists or the IU score for an alternative is judged unacceptable). The decision rule to maximize RIW is easily adapted to situations involving maximum or minimum acceptable values on a given criterion. When such situations arise, simply eliminate alternatives that violate the constraint, regardless of their overall RIW scores.

SECTION 3

CIEA DEVELOPMENTAL ISSUES

Recap

The previous sections have presented an overview of the problem situation addressed by CIEA; i.e., determining the point at which the cost of obtaining the readiness information provided by a DORAC is offset by the information's value to recipients. An approach to assessing information worth based on MAUM was also presented and illustrated with a hypothetical example. The MAUM approach to information worth assessment is illustrated in greater detail in the report Training Device Operational Readiness Assessment Capability (DORAC): Feasibility and Utility (Hawley & Dawdy, 1980). The companion report presents the results of an actual CIEA performed on a set of training devices for the M-16A1 rifle.

As noted in Section 1, IU, the output of the MAUM-based evaluation procedure, is used as a proxy measure for strict IW. The rationale for this approach is based upon the difficulty of assessing IW using other potential methods, such as decision-maker action substitution. Using IU as a proxy measure for IW is based on the assumption that IW is an increasing monotonic function of IU; that is,

$$IW = \Phi(IU), \qquad (3\text{-}1)$$

such that $IW_i \leq IW_j$ whenever $IU_i \leq IU_j$. In other words, as IU increases, IW also increases, or at least does not decrease. Figure 3-1 graphically illustrates one monotonic example of the assumed monotonic relationship expressed in (3-1).

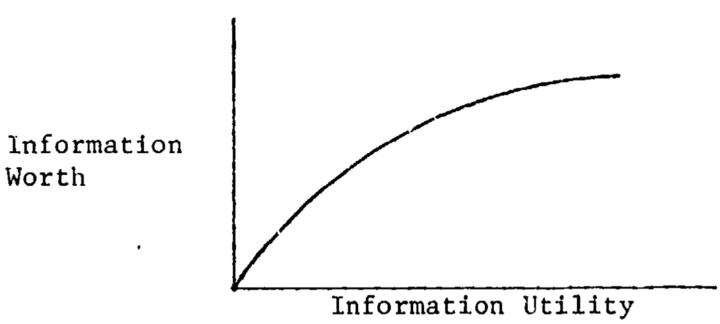

Figure 3-1. Assumed Monotonic Relationship Between Information Worth and Information Utility

An obvious issue to be resolved before a widespread acceptance of the results of MAUM-CIEA is the *validity* of the assumption expressed in (3-1). Establishing the validity of using IU as a proxy measure for IW is basically a two-step procedure. The steps involved in validating the assumption are listed as follows:

1. Establishing that the MAUM-based IU assessment procedure is a well-behaved (i.e., reliable) process. That is, determining whether or not MAUM-derived measures of IU are consistent, equal-interval estimates of IW.

2. Establishing that IW is a monotonically increasing function of IU, as assessed using the MAUM-based assessment procedure. This, essentially, involves establishing the predictive validity of the MAUM-CIEA methodology.

Each of the above issues, plus related methodological issues, is addressed in more detail in the following subsection.

Methodological Issues

The objective of the IU assessment portion of the CIEA methodology is to provide a measure of the "value for decision-making" of a given amount of information obtainable from a DORAC. Stevens (1951) defines measurement as the process of assigning numbers to objects or events according to rules. The MAUM-CIEA methodology outlined in Section 2 defines one set of rules, or one "yardstick", for measuring IW. However, the usability of the MAUM-based yardstick has not been established. To date, little information regarding the actual properties of MAUM-based IU scores has been produced. To be appropriate for use in CIEA, it is necessary to demonstrate that the MAUM procedure is broadly generalizable and that resulting IU scores are:

1. Reliable,
2. Properly scaled (i.e., at least equal-interval), and
3. Predictively valid indices of strict IW.

A primary requirement for any CIEA IW evaluation procedure is that it be broadly generalizable. Within the context of DORAC evaluation, generalizability refers to the methodology being usable with training devices of varying complexity (i.e., ranging from a few performances/conditions with a single measurement capability to a large number of performances/conditions with a sophisticated measurement capability) at various stages in their developmental cycle (e.g., conceptual, bread-board, fielded). The MAUM-CIEA methodology outlined in Section 2 has been demonstrated using a *fielded* training device system having only low to moderate compexity, as indexed by the number of relevant performance objectives and the sophistication of the associated measurement capability (see Hawley & Dawdy, 1980). Hence, the first methodological issue to be addressed concerns establishing that the MAUM-CIEA methodology is generalizable to a range of training devices. This issue can be resolved only through an application of the methodology to a series of training devices spanning the Device Complexity - Developmental Stage axes. Through a series of such applications, implementation and procedural problems can be identified and rectified. The resulting evaluation data can also be used in additional aspects of the validation process, such as investigating reliability, scaling properties, and predictive validity.

Once it has been established that the MAUM-CIEA procedure is operationally generalizable, a second methodological concern is the reliability of the process. In general, reliability refers to the *consistency* from one set of measurements to another on repetition of a measurement procedure (Stanley, 1971). In the case of MAUM-CIEA, reliability denotes the stability or reproducibility of IU results upon repeated application of the methodology by independent sets of raters. An obviously desirable state of affairs is that CIEA results be reasonably independent of whomever constitutes the decision-making group, given that equally qualified representatives of the same stakeholders provide the constituent ratings.

Proceeding from the definition of reliability presented in the previous paragraph, MAUM-CIEA reliability will likely have to be assessed in a manner analogous to that of parallel forms reliability in psychological test theory. In psychological test development, parallel forms reliability is established by first developing two independent testing procedures (i.e., parallel forms) assumed to provide the same true score. Next, each form of the test is administered to equivalent groups of testees. The correlation of results obtained using the two testing procedures provides the basis for computing a reliability coefficient (Lord & Novick, 1968).

In the case of MAUM-CIEA, establishing reliability would be done in a conceptually similar fashion: Independent groups of decision-makers representing the same stakeholders would complete the MAUM procedure evaluating the *same* set of DORAC alternatives. The results of the repeated applications would then be compared at a variety of levels; e.g., importance weights for WDs, partial utility scores for OMs, effectiveness scores for alternatives, and aggregate IU scores. The degree of consistency across levels would provide an indication of the reliability of the MAUM-CIEA methodology. Data used in this empirical examination of reliability could be obtained by replicating certain of the evaluations used to establish generalizability. For example, specified analyses could be conducted on a single set of alternatives using two independent sets of decision-makers.

Strictly speaking, the reliability assessment procedure outlined in the previous paragraph will not demonstrate the absolute reliability of the MAUM method, but instead only the results of its application in a particular situation (e.g., for a training device of a given complexity at a given developmental stage and for a given group of decision-makers) (Torgerson, 1958). Hence, it would be desirable to demonstrate experimentally the reliability of the MAUM-CIEA procedure across a range of devices to which the methodology might be applied. This could be accomplished through a replication of the parallel groups evaluation process across a range of training devices. In any event, even a single demonstration of reliability would serve to enhance user confidence in the resulting data.

A third methodological issue relevant to the acceptability of the MAUM-CIEA IU evaluation procedure concerns the scaling properties of the resulting data. In determining utilities using any of the variations of the Direct scaling methods, it is *assumed* that decision-makers are capable of rating WDs, OMs, and the effectiveness of DORAC alternatives on an *equal-interval* subjective scale. If this assumption is correct, then the procedures presented in Section 2 provide scale values that have equal-interval properties. It should be noted, however, that the scaling methods themselves provide no explicit means of *testing* this assumption.

The assumption that decision-makers are capable of providing equal-interval scale values for WDs, OMs, and system effectiveness is critical to the system evaluation procedures used in MAUM-CIEA. The use of MAUM-derived IU scores in the computation of RIU and RIW is based upon the *assumption* that the level of measurement for IU is *at least* equal-interval (i.e., equal-interval or ratio). The effects on RIU and RIW (and thus the eventual selection of a preferred DORAC alternative) of violations of the equal-interval assumption are not known. A reasonable conclusion, however, is that if IU is at most ordinal, then RIU and RIW are at most ordinal. The use of cost-effectiveness ratios (e.g., RIW) is based on the assumption that both the numerator and denominator terms are at least equal-interval. Using this powerful tool for integrating system cost and effectiveness measures is inappropriate if the scaling properties of either numerator or denominator are suspect. In fact, the undesirable effects of possible violations of the equal-interval assumption have resulted in a general aversion to the use of cost-effectiveness ratios within the military systems analysis community (Paris, 1980).

In view of the above discussion, the validity of the equal-interval assumption should be examined empirically. As in the case of generalizability and reliability, testing the equal-interval assumption would require a repeated application of the MAUM-CIEA procedure in the same evaluation situation. Repeated applications are necessary because data obtained in a single application provide no basis for establishing whether or not decision-makers are judging on the basis of an equal-interval scale (Torgerson, 1958). It is always possible to compute scale values on the basis of an equal-interval assumption. In addition, the consistency of judgments across groups is in itself, not an adequate criterion by which to evaluate the equal-interval assumption. Completely inconsistent judgments are evidence that ratings do not follow an equal-interval scale. Consistent ratings, however, do not imply that decision-makers are judging on an equal-interval scale. A criterion based on consistency alone does not distinguish between equal-interval judgments and straight ordinal position judgments.

The minimum requirement for an equal-interval scale is that the *ratios of differences* in scale values assigned to any three or more stimuli are invariant with respect to the values of the remaining stimuli in the set (Torgerson, 1958). This can be experimentally verified by plotting scale values obtained from one evaluation against the scale values from a second independent replication (easily obtained within the context of replications directed at establishing generalizability and reliability). If the equal-interval assumption is valid, the plot will be linear, within sampling error. Again, as in the case of reliability, a demonstration that decision-makers used an equal-interval scale in one situation will not necessarily generalize to other situations. Repeated demonstrations of the validity of the equal-interval assumption would, however, build user confidence in the applicability of using MAUM-derived IU scores, and thus RIW, in CIEA.

A fourth methodological issue, somewhat related to the issue of the scaling properties of IU scores, concerns overall methodological complexity. The CIEA methodology presented in Section 2 employs a mixture of decomposition and holistic utility evaluation procedures to determine IU. Decisions concerning the use of decomposition versus holistic judgments at various places in the procedure were made on the basis of previous MAUM research and applications and on the basis of *perceived* limits on the complexity of the resulting analytical method. Should the current version of the MAUM-CIEA methodology not provide suitably scaled IU scores, a potential means of raising the level of measurement for IU is to examine the suitability of the utility evaluation procedures used in the analysis. The intent of this examination would be to refine the methodology by using the most appropriate evaluation procedures; that is, by using decomposition methods where they are most appropriate and holistic methods where they are most appropriate. It should be noted, however, that one result of such modifications (e.g., replacing a holistic by a decomposition procedure) could be an increased level of complexity in the analysis and, thus, an elicitation procedure that is more difficult for decision-makers to employ.

The MAUM-CIEA methodology outlined in Section 2 is already relatively complex in terms of process branching, the number of judgments required, and the information load placed on users. As with most MAUM procedures, MAUM-CIEA also tends to be time consuming in its application. Increasing the complexity of the method beyond the present level will act against the widespread applications of MAUM-CIEA as part of the DORAC evaluation process.

One way to alleviate the demands placed on MAUM-CIEA users is to computer-aid the elicitation process. Computer-aiding the elicitation process would involve developing an interactive computer capability to guide and assist decision-makers in applying the methodology. Several studies have indicated that computer aiding MAUM-based decision processes can result in time savings as well as enhanced decision consistency (Freedy, David, Steeb, Samet, & Gardiner, 1976; Sicherman, 1975; Ulvila, 1975).

A fifth methodological issue relevant to the application of MAUM-CIEA concerns the predictive validity of IU scores; that is, establishing that IU can, in fact, be used as a proxy measure for strict IW. As noted previously, IU is used as a proxy measure for IW under the *assumption* that IW is an increasing monotonic function of IU. Hence, establishing that IU can be used in lieu of strict IW reduces to the problem of validating the assumption of monotonicity.

Given the current status of CIEA methodological development, the assumption of monotonicity will likely have to be tested within a convergent validation framework (Campbell & Fiske, 1959). In the present situation, convergent validation would involve determining the extent to which IU (obtained using MAUM) is related to *other independently derived* measures of IW. Hence, the first requirement in validating the use of IU as a proxy measure for strict IW involves developing one or more independent methods for assessing IW.

At the present time, IW is rather loosely defined as "value for decision-making". Before developing alternative methods for assessing IW, it will first be necessary to define IW in exact terms. Carnap (1950) has referred to this process as *explication*, or operational definition. Along with the operational definition, rules for the measurement of the construct are also specified. Then, given an operational definition for IW and rules for its measurement, the next step in validating MAUM-CIEA is to assess IW using one or more independent procedures. Agreement between IU, as obtained using MAUM-CIEA, and the other independent measures of IW would lend support to the use of IU as a proxy measure for IW.

One approach that may hold promise for developing an alternative measure of IW involves the use of combat simulation models (CSMs) (e.g., ASARS, COTEAM, CARMONETTE, etc.). Conceptually speaking, CSMs have the potential for determining the nature of the relationship between certain aspects of individual or collective performance (e.g., M-16 A1 accuracy) and overall individual or unit combat effectiveness. That is, CSMs may have potential for specifying (at least partly) the nature of the production function expressed as (3-2):

$$\underline{\varepsilon} = f(x_1, x_2, x_3, \ldots, x_n), \tag{3-2}$$

where $\underline{\varepsilon}$ represents individual or unit combat effectiveness,

the x_i represent the *level* of performance on component $\underline{i}$,

and $f(\cdot)$ is a function relating performance levels to $\underline{\varepsilon}$.

Should it prove possible to form (3-2), it then reasonably follows that IW could be defined as a function of the parameters relating performance levels to $\underline{\varepsilon}$. For example, if changes in M-16 A1 accuracy do not affect $\underline{\varepsilon}$, then information concerning M-16 A1 accuracy has little real value; whereas, in the same situation, if fire distribution patterns are found to impact significantly on $\underline{\varepsilon}$, then information concerning fire distribution patterns has high value. Mathematically, the relationship between the parameters relating the x_i to $\underline{\varepsilon}$ and IW can be expressed as follows:

$$IW = g(\underline{\alpha}_1, \underline{\alpha}_2, \underline{\alpha}_3, \ldots, \underline{\alpha}_n), \tag{3-3}$$

where IW represents information worth,

the $\underline{\alpha}_i$ represent the parameters in $f(\cdot)$ relating the x_i to $\underline{\varepsilon}$,

and $g(\cdot)$ is a function relating the $\underline{\alpha}_i$ to IW.

Short of assessing results of actual (or near actual) combat situations, attempting to evaluate before and after decision scenarios (see the discussion presented in Section 1), or arbitrarily defining IU to be an acceptable measure of strict IW, the CSM approach to developing an alternate measure of IW is the only viable means for establishing the predictive validity of MAUM-CIEA. Relying on user opinions regarding the reasonableness (i.e., face validity) of MAUM-CIEA results is not sufficient. Face validity can be unrelated to predictive or criterion-referenced validity (Lord & Novick, 1968). The MAUM-CIEA methodology provides a means for assessing the perceptual set of a group of decision-makers. If the perceptual set is in error, then all an additional round of subjective evaluations will accomplish is to validate existing institutional prejudices. As with other evaluative methodologies, face validity is important to the acceptance of results. In the final analysis, however, the results of MAUM-CIEA should be exposed at least once to the acid test of objective reality, even if that reality is determined in a somewhat contrived or limited situation.

The previous points should not be taken to imply that CSM-CIEA (or a variant) is always to be preferred to MAUM-CIEA. Given that IW is demonstrated to be monotonically related to IU, MAUM-CIEA will be preferred in most situations. For example, appropriate CSMs may not exist or the cost of CSM-CIEA might preclude its practical application. It should be emphasized that the application of CSM-CIEA is likely to be quite expensive relative to an application of MAUM-CIEA. The point remains, however, that the predictive validity of MAUM-CIEA is a critical issue and should be empirically evaluated, to the extent deemed practical and sufficient.

An ancillary result of the successful development of a CSM-CIEA methodology would be the potential for specifying more *veridical* performance standards for the performances relevant to the DORAC under study. Veridical performance standards are defined as those that are based upon the true contribution of individual performances to combat effectiveness. If the function expressed as (3-2) can be specified, then a potential exists for establishing realistic standards for the performances involved in the x_i. As in personnel selection, the first step in such a developmental process would be to specify a minimally acceptable level for $\underline{\varepsilon}$. Then, the minimum levels for various x_i required to produce the minimum level of $\underline{\varepsilon}$ could directly be determined. Interactions, or trade offs, between levels of x_i and $\underline{\varepsilon}$ could also be studied. The resulting data could prove useful in the conduct of CTEA and other training development or evaluation activities.

In closing, this report has presented a methodology, MAUM-CIEA, for establishing the cost-effectiveness of information provided by potential DORACs. As noted earlier in this section of the report, several methodological issues relevant to MAUM-CIEA remain points of inquiry. This situation is, however, not unusual in the case of newly developed analytical procedures, even those that are integrated from proven components. The issues noted in this section should be explored. Exploring them will take time, but in the interim MAUM-CIEA need not remain on the shelf. If MAUM-CIEA results are viewed cautiously until more data regarding the methodological issues are available, the methodology can (and should) be used in the evaluation of potential DORACs or in other potentially-relevant aspects of the training device development process. Data obtained from such initial applications of MAUM-CIEA could prove highly useful in streamlining the procedure and in making the method more responsive to actual user needs.

References

American Airlines, Inc. Optimized flight crew training: A step toward safer operations. Ft. Worth, Texas: American Airlines Flight Training Academy, April 1969.

Barish, N.N., & Kaplan, S. Economic analysis for engineering and managerial decision making (2nd ed.). New York: McGraw-Hill, 1978.

Bedford, N.M., & Onsi, M. Measuring the value of information--An information theory approach. Management Services, January-February, 1966, 1, 15-22.

Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Carnap, R. Logical foundations of probability. Chicago: University of Chicago Press, 1950.

Caro, P. Aircraft simulators and pilot training. Human Factors, 1973, 15(6), 502-509.

Churchman, C.W., Ackoff, R.L., & Arnoff, E.L. Introduction to operations research. New York: John Wiley, 1957.

Coombs, C.H. A theory of data. New York: John Wiley, 1964.

Crawford, A., & Brock, J. Using simulators for performance measurement. Proceedings of the Symposium on Productivity Enhancement: Personnel Assessment in Navy Systems, San Diego, California: Naval Personnel Research and Development Center, October 1977.

Dawes, R.M. A case study of graduate admissions: Applications of three principles of human decision making. American Psychologist, 1971, 26, 180-188.

Dawes, R.M., & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-105.

Edwards, W. Social utilities. The Engineering Economist, Summer Symposium Series, 1971, 6, 119-123.

Edwards, W. How to use multi-attribute utility measurement for social decision making (SSRI Tech. Rep. 76-3-1-T). University of Southern California, Los Angeles, California: Social Science Research Institute, August 1976.

Einhorn, H.J., & Hogarth, R.M. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Finley, D.L., Gainer, C.A., & Muckler, F.A. Test and evaluation of a carrier air traffic control center team trainer as a performance qualification instrument: Phase I report. Canoga Park, California: XYZYX Information Corp., 1974. (NTIS No. AD B003 213L)

Fitzpatrick, R., & Morrison, E.J. Performance and product evaluation. In R.L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Freedy, A., David, K.B., Steeb, R., Samet, M.G., & Gardiner, P.C. Adaptive computer aiding in dynamic decision processes: Methodology, evaluation and application (Technical Report). Woodland Hills, Calif.: Perceptronics Inc., 1976.

Glaser, R., & Klaus, D.J. Proficiency measurement: Assessing human performance. In R.M. Gagne (Ed.), Psychological principles in system development. New York: Holt, Rinehart, and Winston, 1972.

Goldberg, L.R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. Psychological Monographs, 1965, 79, (9, Whole No. 602).

Goldberg, L.R. Simple models or simple processes? Some research on clinical judgment. American Psychologist, 1968, 23, 483-496.

Goldberg, L.R. Man versus model of man: A rationale plus some evidence for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.

Goldberg, L.R. Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. Organizational Behavior and Human Performance, 1971, 6, 458-479.

Guilford, J.P. Psychometric methods. New York: McGraw-Hill, 1954.

Hawley, J.K., & Dawdy, E.D. Training device operational readiness assessment capability (DORAC): Feasibility and utility (Technical Report). Valencia, Pa.: Applied Science Associates, April 1981.

Hopkins, C.O. How much should you pay for that box? In Proceedings of the 19th Annual Meeting of the Human Factors Society. Dallas, Tex.: Human Factors Society, 1975, i-vi.

John, R.S., & Edwards, W. Importance weight assessment for additive, riskless preference functions: A review (SSRI Research Report 78-5). Los Angeles, Calif.: Social Science Research Institute, December 1978.

Johnson, E.M., & Huber, P.G. The technology of utility assessment. IEEE Transactions on Systems, Man, and Cybernetics, 1977, SMC-7, 311-325.

Kazanowski, A.D. A standardized approach to cost effectiveness evaluation. In: J.M. English (Ed.), Cost-Effectiveness. New York: John Wiley, 1968.

Keeney, R.L. Utility functions for multi-attributed consequences. Management Science, 1976, 18, 276-287.

Keeney, R.L. The art of assessing multi-attribute utility functions. Organizational Behavior and Human Performance, 1977, 19, 267-310.

Keeney, R.L. & Raiffa, H. Decisions with multiple objectives: Preferences and value trade-offs. New York: John Wiley, 1976.

Lev, B. Accounting and information theory. Chicago: American Accounting Association, 1969.

Lord, F.M., & Novick, M.R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Paris, L. Personal communication, October 15, 1980.

Raiffa, H. Decision analysis. Reading, Mass.: Addison-Wesley, 1968.

Shannon, C.E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27, pp. 379-423; 623-656.

Shannon, C.E. Systems simulation: The art and science. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.

Shannon, C.E., & Weaver, W. The mathematical theory of communication. Urbana, Ill.: University of Illinois Press, 1963.

Shelnutt, J.B., Smillie, R.J., & Bercos, J. A consideration of Army training device proficiency assessment capabilities (ARI Tech. Rep. TR-78-A20). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, June 1978.

Sicherman, A. An interactive computer program for assessing and using multiattribute utility functions (Technical Report). Cambridge, Mass.: Massachusetts Institute of Technology, Cambridge Operations Research Center, 1975.

Stanley, J.C. Reliability. In R.L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1971.

Stevens, S.S. Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), Handbook of Experimental Psychology. New York: John Wiley, 1951.

Stevens, S.S. Psychophysics. New York: John Wiley, 1975.

Swezey, R.W. & Pearlstein, R.B. Developing criterion referenced tests (Technical Report). Valencia, Pa.: Applied Science Associates, Inc., December 1974.

Thiel, H. Economics and information theory. Chicago: Rand McNally and North Holland Publishing Co., 1967.

Torgerson, W.S. Theory and methods of scaling. New York: John Wiley, 1958.

Ulvila, J.W. A pilot survey of computer programs for decision analysis (Technical Report). McLean, Va.: Decisions and Designs, Inc., 1975.

U.S. Army Training and Doctrine Command. Cost and operational effectiveness handbook (TRADOC PAM 11-8). Ft. Monroe, Va.: Author, November 1974.

Weitzman, D.O., Fineberg, M.L., Gade, P.A., & Compton, G.L. Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 1979, 21(6), 701-710.

Yntema, D.B., & Torgerson, W.S. Man-computer cooperation in decisions requiring common sense. IRE Transactions of the Professional Group on Human Factors in Electronics, 1961, HFE-2, 20-26.

APPENDIX A

DORAC COSTING GUIDELINE

Introduction

The objective of the following costing guideline is to lead an analyst through a series of steps that will lead to an estimate of the cost of developing, acquiring, and operating a DORAC. Factors that are considered in estimating the total cost of the capability include the following:

1. Facilities--Physical assets (buildings, land, etc.) required for the implementation and support of a DORAC system.

2. Equipment--Hardware directly associated with evaluation positions and their support.

3. Materials--Software (user materials, films, texts, computer programs, etc.) necessary for the conduct and support of DORAC evaluations.

4. Personnel--Evaluator and evaluatee costs incurred as a *direct* result of DORAC evaluation activities (e.g., salaries, travel, per diem, expendable supplies, etc.).

The first step in costing a DORAC is to specify the exact nature of each alternative; that is, to list the devices required, the planned number of evaluations per year, and support requirements in terms of operators, evaluators, and maintenance personnel; and so forth. Then, for each training device comprising part of an alternative and for the baseline alternative, the following series of steps is completed:

Step 1. Compute Anticipated Facility Load (AFL)

a. Specify Annual Evaluation Frequency (AEF), the number of times each evaluatee must be evaluated annually using the device.

b. For the device, specify the Evaluation Cycle Size (ECS), the number of evaluation positions that can be used simultaneously.

c. Identify Organization Evaluation Unit Size (OEUS), the evaluatee group size (e.g., individual, squad, platoon, company, etc.) organic to the base organization (e.g., battalion, brigade, division, post, etc.) that most closely matches, without exceeding, the ECS.

d. Specify the number of Evaluatees per Organization (E/O), the total number of personnel within the base organization that the evaluation is applicable to.

e. Compute Organizational Evaluations (OE), the number of evaluation cycles required to process one base organization:

$$OE = \frac{E/O}{OEUS} . \tag{A-1}$$

f. Compute Base Facility Load (BFL), the number of evaluations required to put all relevant base organizations (N_{org}) through one evaluation cycle:

$$BFL = N_{org} \times OE, \tag{A-2}$$

where N_{org} is the number of base organizations scheduled to undergo evaluation.

g. Compute Estimated Facility Load (EFL), the expected number of evaluation cycles to process all base organizations on all evaluations:

$$EFL = BFL \times AEF. \tag{A-3}$$

h. Determine the expected number of re-evaluation (RE) cycles:

$$RE = PLE \times EFL, \tag{A-4}$$

where PLE is the estimated percentage of lost evaluations.

i. Finally, compute AFL, the expected number of evaluation cycles required to process all evaluatees, plus expected re-evaluations:

$$AFL = RE + EFL. \tag{A-5}$$

<u>Step 2. Compute Device Capability (DC)</u>

a. Determine Work Segment Length (WSL), the length of time in hours between work starts and stops during the work day.

b. Specify the Number of Work Segments (NWS), the number of WSL periods per work day.

c. Estimate Evaluation Cycle Length (ECL), the anticipated time in hours required to conduct one evaluation cycle using the device.

d. Estimate Time Between Cycles (TBC), time in hours spent in orientation briefing, data collection, or critique.

e. Compute the number of Evaluation Cycles per Day (EC/D) (rounded to nearest integer):

$$EC/D = \frac{WSL \times SN}{ECL + TBC} . \tag{A-6}$$

f. Estimate Time Device Available (TDA), the number of training days per year the device is expected to be operational (round to nearest day):

$$TDA = PTA \times TDY, \tag{A-7}$$

where PTA is the percentage of time the device is expected to be available,

and TDY is the number of training days per year.

g. Compute Device Capability (DC), the expected number of evaluation cycles per year that one device will accommodate:

$$DC = TDA \times (EC/D). \tag{A-8}$$

Step 3. Compute the Number of Devices Required (NDR):

$$NDR = AFL/DC. \tag{A-9}$$

Step 4. Determine Incremental Acquisition Cost of Device (ACD)

a. Compute Device Hardware Cost (DHC), the cost of acquiring the required number of devices (above the number required solely for training):

$$DHC = (NDR - NDOH) \times UCD, \tag{A-10}$$

where NDOH is the number of devices on hand for training purposes,

and UCD is the unit cost of the device.

b. Determine Direct Personnel Costs (DPC). DPC is the sum of the estimated number by proposed pay grade of operators, evaluators, and maintenance personnel multiplied by their salaries and benefits. Again, this is the incremental cost associated with providing personnel for purely evaluation purposes.

c. Determine evaluation-related Personnel Training Costs (PTC):

$$PTC = CIT + EFT, \tag{A-11}$$

where CIT is the cost of initial training for operators, evaluators, and maintainers;

and EFT is the expected cost of follow-on training (if any) plus the cost of training replacements for the initial cadre.

d. Compute Administrative Operating Costs (AOC):

$$AOC = EXS + TC, \tag{A-12}$$

where EXS is the cost of data recording supplies, office supplies, facilities supplies, equipment supplies, utilities, ammunition, targets, and so forth associated with evaluation activities,

and TC is the cost of any required travel for evaluatees or cadre (again, evaluation related).

e. Estimate Facilities Cost (FC). FC is the total cost of required acquisition, construction, or modification for device-related facilities used solely for evaluation.

f. Estimate Annual Maintenance Cost (AMC), the annual cost of maintaining device-related facilities used solely for evaluation.

g. Compute ACD:

$$ACD = DHC + (DPC + AOC + AMC) \times EY + PTC + FC, \tag{A-13}$$

where EY is the expected life of the device.

Step 5. Determine Total Device Cost (TDC):

$$TDC = ACD + EEIR \times DHC \times EY, \qquad \text{(A-14)}$$

where EEIR is the expected replacement rate (percentage) per year for end item equipment used for evaluation purposes.

Step 6. Determine Total Estimated Cost of Alternative (TECA):

After the estimated cost of each device constituting a DORAC alternative is calculated, these costs are summed to produce the Total Estimated Cost of the Alternative (TECA):

$$TECA = \sum_{i=1}^{NDEV} TDC_i, \qquad \text{(A-15)}$$

where NDEV is the number of individual devices constituting DORAC alternative i,

and TDC_i is the estimated cost of the i^{th} device (from A-14).

The application of (A-15) in determining the total estimated cost of alternative i makes the assumption that all individual device costs are mutually exclusive; that is, no shared costs (e.g., common operators, evaluators, facilities, etc.) are involved. In the event that shared costs are involved, assign the overlapping costs to one device. In this manner, shared costs are only counted one time.